WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

A61K

(11) International Publication Number: WO 99/42076

(43) International Publication Date: 26 August 1999 (26.08.99)

(21) International Application Number: PCT/US99/03268

(22) International Filing Date: 17 February 1999 (17.02.99)

(30) Priority Data:

09/025,197 18 February 1998 (18.02.98) US 09/072,967 5 May 1998 (05.05.98) US

(71) Applicant: CORIXA CORPORATION [US/US]; 1124 Columbia Street, Seattle, WA 98104 (US).

(72) Inventors: REED, Steven, G.; 2843 – 122nd Place, N.E., Bellevue, WA 98005 (US). SKEIKY, Yasir, A., W.; 8327 25th Street, Seattle, WA 98107 (US). DILLON, Davin, C.; 21607 N.E. 24th Street, Redmond, WA 98053 (US). CAMPOS-NETO, Antonio; 9308 Midship Court, N.E., Bainbridge Island, WA 98110 (US). HOUGHTON, Raymond; 2636 – 242nd Place, S.E., Bothell, WA 98021 (US). VEDVICK, Thomas, S.; 124 South 300th Place, Federal Way, WA 98003 (US). TWARDZIK, Daniel, R.; 10195 South Beach Drive, Bainbridge Island, WA 98110 (US). LODES, Michael, J.; 9223 – 36th Avenue S.W., Seattle, WA 98126 (US). HENDRICKSON, Ronald, C.; 4114 S.W. Charlestown Street, Seattle, WA 98116 (US).

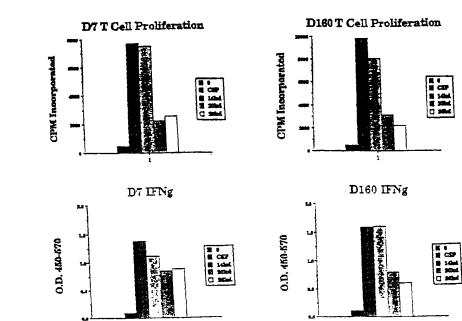
(74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS



(57) Abstract

Compounds and methods for inducing protective immunity against tuberculosis are disclosed. The compounds provided include polypeptides that contain at least one immunogenic portion of one or more *M. tuberculosis* proteins and DNA molecules encoding such polypeptides. Such compounds may be formulated into vaccines and/or pharmaceutical compositions for immunization against *M. tuberculosis* infection, or may be used for the diagnosis of tuberculosis.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

ΑL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal .
ΑU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
ΑZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	ΙE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	ΙT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL.	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

WO 99/42076

1

COMPOUNDS AND METHODS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS

5 CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Application No. 9/025,197, filed February 18, 1998; which is a continuation-in-part of U.S. Application No. 08/942,578, filed October 1, 1997; which is a continuation-in-part of U.S. Application No. 08/818,112, filed March 13, 1997; which is a continuation-in-part of U.S. Application No. 08/730.510, filed October 11, 1996; which claims priority from PCT Application No. PCT/US 96/14674, filed August 30, 1996; and is a continuation-in-part of U.S. Application No. 08/680,574, filed July 12, 1996; which is a continuation-in-part of U.S. Application No. 08/659,683, filed June 5, 1996; which is a continuation-in-part of U.S. Application No. 08/620,874, filed March 22, 1996, now abandoned; which is a continuation-in-part of U.S. Application No. 08/533.634, filed September 22, 1995, now abandoned; which is a continuation-in-part of U.S. Application No. 08/523,436, filed September 1, 1995, now abandoned.

TECHNICAL FIELD

The present invention relates generally to detecting, treating and preventing Mycobacterium tuberculosis infection. The invention is more particularly related to polypeptides comprising a Mycobacterium tuberculosis antigen, or a portion or other variant thereof, and the use of such polypeptides for diagnosing and vaccinating against Mycobacterium tuberculosis infection.

25

30

20

10

15

BACKGROUND OF THE INVENTION

Tuberculosis is a chronic, infectious disease, that is generally caused by infection with *Mycobacterium tuberculosis*. It is a major disease in developing countries, as well as an increasing problem in developed areas of the world, with about 8 million new cases and 3 million deaths each year. Although the infection may be asymptomatic for a considerable period of time, the disease is most commonly

2

manifested as an acute inflammation of the lungs, resulting in fever and a nonproductive cough. If left untreated, serious complications and death typically result.

Although tuberculosis can generally be controlled using extended antibiotic therapy, such treatment is not sufficient to prevent the spread of the disease. Infected individuals may be asymptomatic, but contagious, for some time. In addition, although compliance with the treatment regimen is critical, patient behavior is difficult to monitor. Some patients do not complete the course of treatment, which can lead to ineffective treatment and the development of drug resistance.

Inhibiting the spread of tuberculosis requires effective vaccination and accurate, early diagnosis of the disease. Currently, vaccination with live bacteria is the most efficient method for inducing protective immunity. The most common Mycobacterium employed for this purpose is *Bacillus* Calmette-Guerin (BCG), an avirulent strain of *Mycobacterium bovis*. However, the safety and efficacy of BCG is a source of controversy and some countries, such as the United States, do not vaccinate the general public. Diagnosis is commonly achieved using a skin test, which involves intradermal exposure to tuberculin PPD (protein-purified derivative). Antigen-specific T cell responses result in measurable induration at the injection site by 48-72 hours after injection, which indicates exposure to Mycobacterial antigens. Sensitivity and specificity have, however, been a problem with this test, and individuals vaccinated with BCG cannot be distinguished from infected individuals.

10

15

20

30

While macrophages have been shown to act as the principal effectors of *M tuberculosis* immunity. T cells are the predominant inducers of such immunity. The essential role of T cells in protection against *M. tuberculosis* infection is illustrated by the frequent occurrence of *M. tuberculosis* in AIDS patients, due to the depletion of CD4 T cells associated with human immunodeficiency virus (HIV) infection. Mycobacterium-reactive CD4 T cells have been shown to be potent producers of gamma-interferon (IFN-γ), which, in turn, has been shown to trigger the antimycobacterial effects of macrophages in mice. While the role of IFN-γ in humans is less clear, studies have shown that 1.25-dihydroxy-vitamin D3, either alone or in combination with IFN-γ or tumor necrosis factor-alpha, activates human macrophages

to inhibit *M. tuberculosis* infection. Furthermore, it is known that IFN- γ stimulates human macrophages to make 1,25-dihydroxy-vitamin D3. Similarly, IL-12 has been shown to play a role in stimulating resistance to *M. tuberculosis* infection. For a review of the immunology of *M. tuberculosis* infection see Chan and Kaufmann in *Tuberculosis: Pathogenesis, Protection and Control*, Bloom (ed.), ASM Press, Washington, DC, 1994.

Accordingly, there is a need in the art for improved vaccines and methods for preventing, treating and detecting tuberculosis. The present invention fulfills these needs and further provides other related advantages.

10

20

25

SUMMARY OF THE INVENTION

Briefly stated, this invention provides compounds and methods for preventing and diagnosing tuberculosis. In one aspect, polypeptides are provided comprising an immunogenic portion of a soluble *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications. In one embodiment of this aspect, the soluble antigen has one of the following N-terminal sequences:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu: (SEQ ID No. 120)
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser: (SEQ ID No. 121)
 - (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 122)
 - (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 123)
 - (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val: (SEQ ID No. 124)
 - (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID No. 125)
- 30 (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser: (SEQ ID No. 126)

4

(h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 127)

- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn; (SEQ ID No. 128)
- (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 134)
- (k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID No. 135) or
- (l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136)

wherein Xaa may be any amino acid.

5

10

20

30

In a related aspect, polypeptides are provided comprising an immunogenic portion of an *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications, the antigen having one of the following N-terminal sequences:

- (m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val: (SEQ ID No. 137) or
- (n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe: (SEQ ID No. 129)

wherein Xaa may be any amino acid.

In another embodiment, the soluble *M. tuberculosis* antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 1, 2, 4-10, 13-25, 52, 99 and 101, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 1, 2, 4-10, 13-25, 52, 99 and 101 or a complement thereof under moderately stringent conditions.

In a related aspect, the polypeptides comprise an immunogenic portion of a *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications, wherein the antigen comprises an

25

amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 26-51, 138, 139, 163-183, 201, 240, 242-247, 253-256, 295-298, 309, 316, 318-320, 322, 324, 328, 339, 333, 335, 337, 339 and 341, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 26-51, 138, 139, 163-183, 201, 240, 242-247, 253-256, 295-298, 309, 316, 318-320, 322, 324, 328, 329, 333, 335, 337, 339 and 341 or a complement thereof under moderately stringent conditions.

In related aspects, DNA sequences encoding the above polypeptides, expression vectors comprising these DNA sequences and host cells transformed or transfected with such expression vectors are also provided.

In another aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, an inventive polypeptide and a known *M. tuberculosis* antigen.

Within other aspects, the present invention provides pharmaceutical compositions that comprise one or more of the above polypeptides, or a DNA molecule encoding such polypeptides, and a physiologically acceptable carrier. The invention also provides vaccines comprising one or more of the polypeptides as described above and a non-specific immune response enhancer, together with vaccines comprising one or more DNA sequences encoding such polypeptides and a non-specific immune response enhancer.

In yet another aspect, methods are provided for inducing protective immunity in a patient, comprising administering to a patient an effective amount of one or more of the above polypeptides.

In further aspects of this invention, methods and diagnostic kits are provided for detecting tuberculosis in a patient. The methods comprise contacting dermal cells of a patient with one or more of the above polypeptides and detecting an immune response on the patient's skin. The diagnostic kits comprise one or more of the above polypeptides in combination with an apparatus sufficient to contact the polypeptide with the dermal cells of a patient.

20

In yet other aspects, methods are provided for detecting tuberculosis in a patient, such methods comprising contacting dermal cells of a patient with one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and detecting an immune response on the patient's skin. Diagnostic kits for use in such methods are also provided.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE IDENTIFIERS

Figure 1A and B illustrate the stimulation of proliferation and interferonproduction in T cells derived from a first and a second *M. tuberculosis*-immune donor. respectively, by the 14 Kd. 20 Kd and 26 Kd antigens described in Example 1.

Figure 2 illustrates the stimulation of proliferation and interferon- γ production in T cells derived from an *M. tuberculosis*-immune individual by the two representative polypeptides TbRa3 and TbRa9.

Figures 3A-D illustrate the reactivity of antisera raised against secretory

M. tuberculosis proteins, the known M. tuberculosis antigen 85b and the inventive antigens Tb38-1 and TbH-9, respectively, with M. tuberculosis lysate (lane 2), M. tuberculosis secretory proteins (lane 3), recombinant Tb38-1 (lane 4), recombinant TbH-9 (lane 5) and recombinant 85b (lane 5).

Figure 4A illustrates the stimulation of proliferation in a TbH-9-specific T cell clone by secretory *M. tuberculosis* proteins, recombinant TbH-9 and a control antigen. TbRall.

Figure 4B illustrates the stimulation of interferon-y production in a TbH-9-specific T cell clone by secretory M. tuberculosis proteins, PPD and recombinant TbH-9.

Figures 5A and B illustrate the stimulation of proliferation and interferon-y production in TbH9-specific T cells by the fusion protein TbH9-Tb38-1.

Figures 6A and B illustrate the stimulation of proliferation and interferon-γ production in Tb38-1-specific T cells by the fusion protein TbH9-Tb38-1.

Figures 7A and B illustrate the stimulation of proliferation and interferon-y production in T cells previously shown to respond to both TbH-9 and Tb38-1 by the fusion protein TbH9-Tb38-1.

Figures 8A and B illustrate the stimulation of proliferation and interferon-γ production in T cells derived from a first M. tuberculosis-immune individual by the representative polypeptides XP-1, RDIF6, RDIF8, RDIF10 and RDIF11.

Figures 9A and B illustrate the stimulation of proliferation and interferon-γ production in T cells derived from a second *M. tuberculosis*-immune individual by the representative polypeptides XP-1, RDIF6, RDIF8, RDIF10 and RDIF11.

SEQ. ID NO. 1 is the DNA sequence of TbRa1.

SEQ. ID NO. 2 is the DNA sequence of TbRa10.

SEQ. ID NO. 3 is the DNA sequence of TbRall.

SEQ. ID NO. 4 is the DNA sequence of TbRa12.

SEQ. ID NO. 5 is the DNA sequence of TbRa13.

SEQ. ID NO. 6 is the DNA sequence of TbRa16.

SEQ. ID NO. 7 is the DNA sequence of TbRa17.

25

SEQ. ID NO. 8 is the DNA sequence of TbRa18.

SEQ. ID NO. 9 is the DNA sequence of TbRa19.

SEQ. ID NO. 10 is the DNA sequence of TbRa24.

SEQ. ID NO. 11 is the DNA sequence of TbRa26.

SEQ. ID NO. 12 is the DNA sequence of TbRa28.

	SEQ. ID NO. 13 is the DNA sequence of TbRa29.
	SEQ. ID NO. 14 is the DNA sequence of TbRa2A.
	SEQ. ID NO. 15 is the DNA sequence of TbRa3.
	SEQ. ID NO. 16 is the DNA sequence of TbRa32.
5	SEQ. ID NO. 17 is the DNA sequence of TbRa35.
	SEQ. ID NO. 18 is the DNA sequence of TbRa36.
	SEQ. ID NO. 19 is the DNA sequence of TbRa4.
	SEQ. ID NO. 20 is the DNA sequence of TbRa9.
	SEQ. ID NO. 21 is the DNA sequence of TbRaB.
10	SEQ. ID NO. 22 is the DNA sequence of TbRaC.
	SEQ. ID NO. 23 is the DNA sequence of TbRaD.
	SEQ. ID NO. 24 is the DNA sequence of YYWCPG.
	SEQ. ID NO. 25 is the DNA sequence of AAMK.
	SEQ. ID NO. 26 is the DNA sequence of TbL-23.
15	SEQ. ID NO. 27 is the DNA sequence of TbL-24.
	SEQ. ID NO. 28 is the DNA sequence of TbL-25.
	SEQ. ID NO. 29 is the DNA sequence of TbL-28.
	SEQ. ID NO. 30 is the DNA sequence of TbL-29.
	SEQ. ID NO. 31 is the DNA sequence of TbH-5.
20	SEQ. ID NO. 32 is the DNA sequence of TbH-8.
	SEQ. ID NO. 33 is the DNA sequence of TbH-9.
	SEQ. ID NO. 34 is the DNA sequence of TbM-1.
	SEQ. ID NO. 35 is the DNA sequence of TbM-3.
	SEQ. ID NO. 36 is the DNA sequence of TbM-6.
25	SEQ. ID NO. 37 is the DNA sequence of TbM-7.
	SEQ. ID NO. 38 is the DNA sequence of TbM-9.
	SEQ. ID NO. 39 is the DNA sequence of TbM-12.
	SEQ. ID NO. 40 is the DNA sequence of TbM-13.
	SEQ. ID NO. 41 is the DNA sequence of TbM-14.
30	SEQ. ID NO. 42 is the DNA sequence of TbM-15.

15

25

SEQ. ID NO. 43 is the DNA sequence of TbH-4.

SEQ. ID NO. 44 is the DNA sequence of TbH-4-FWD.

SEQ. ID NO. 45 is the DNA sequence of TbH-12.

SEQ. ID NO. 46 is the DNA sequence of Tb38-1.

SEQ. ID NO. 47 is the DNA sequence of Tb38-4.

SEQ. ID NO. 48 is the DNA sequence of TbL-17.

SEQ. ID NO. 49 is the DNA sequence of TbL-20.

SEQ. ID NO. 50 is the DNA sequence of TbL-21.

SEQ. ID NO. 51 is the DNA sequence of TbH-16.

SEQ. ID NO. 52 is the DNA sequence of DPEP.

SEQ. ID NO. 53 is the deduced amino acid sequence of DPEP.

SEQ. ID NO. 54 is the protein sequence of DPV N-terminal Antigen.

SEQ. ID NO. 55 is the protein sequence of AVGS N-terminal Antigen.

SEQ. ID NO. 56 is the protein sequence of AAMK N-terminal Antigen.

SEQ. ID NO. 57 is the protein sequence of YYWC N-terminal Antigen.

SEQ. ID NO. 58 is the protein sequence of DIGS N-terminal Antigen.

SEQ. ID NO. 59 is the protein sequence of AEES N-terminal Antigen.

SEQ. ID NO. 60 is the protein sequence of DPEP N-terminal Antigen.

SEQ. ID NO. 61 is the protein sequence of APKT N-terminal Antigen.

SEQ. ID NO. 62 is the protein sequence of DPAS N-terminal Antigen.

SEQ. ID NO. 63 is the deduced amino acid sequence of TbRa1.

SEQ. ID NO. 64 is the deduced amino acid sequence of TbRa10.

SEQ. ID NO. 65 is the deduced amino acid sequence of TbRall.

SEQ. ID NO. 66 is the deduced amino acid sequence of TbRa12.

SEQ. ID NO. 67 is the deduced amino acid sequence of TbRa13.

SEQ. ID NO. 68 is the deduced amino acid sequence of TbRa16.

SEQ. ID NO. 69 is the deduced amino acid sequence of TbRa17.

SEQ. ID NO. 70 is the deduced amino acid sequence of TbRa18.

SEQ. ID NO. 71 is the deduced amino acid sequence of TbRa19.

SEQ. ID NO. 72 is the deduced amino acid sequence of TbRa24.

25

- SEQ. ID NO. 73 is the deduced amino acid sequence of TbRa26.
- SEQ. ID NO. 74 is the deduced amino acid sequence of TbRa28.
- SEQ. ID NO. 75 is the deduced amino acid sequence of TbRa29.
- SEQ. ID NO. 76 is the deduced amino acid sequence of TbRa2A.
- SEQ. ID NO. 77 is the deduced amino acid sequence of TbRa3.
 - SEQ. ID NO. 78 is the deduced amino acid sequence of TbRa32.
 - SEQ. ID NO. 79 is the deduced amino acid sequence of TbRa35.
 - SEQ. ID NO. 80 is the deduced amino acid sequence of TbRa36.
 - SEQ. ID NO. 81 is the deduced amino acid sequence of TbRa4.
- SEQ. ID NO. 82 is the deduced amino acid sequence of TbRa9.
 - SEQ. ID NO. 83 is the deduced amino acid sequence of TbRaB.
 - SEQ. ID NO. 84 is the deduced amino acid sequence of TbRaC.
 - SEQ. ID NO. 85 is the deduced amino acid sequence of TbRaD.
 - SEQ. ID NO. 86 is the deduced amino acid sequence of YYWCPG.
- SEQ. ID NO. 87 is the deduced amino acid sequence of TbAAMK.
 - SEQ. ID NO. 88 is the deduced amino acid sequence of Tb38-1.
 - SEQ. ID NO. 89 is the deduced amino acid sequence of TbH-4.
 - SEQ. ID NO. 90 is the deduced amino acid sequence of TbH-8.
 - SEQ. ID NO. 91 is the deduced amino acid sequence of TbH-9.
 - SEQ. ID NO. 92 is the deduced amino acid sequence of TbH-12.
 - SEQ. ID NO. 93 is the amino acid sequence of Tb38-1 Peptide 1.
 - SEQ. ID NO. 94 is the amino acid sequence of Tb38-1 Peptide 2.
 - SEQ. ID NO. 95 is the amino acid sequence of Tb38-1 Peptide 3.
 - SEQ. ID NO. 96 is the amino acid sequence of Tb38-1 Peptide 4.
 - SEQ. ID NO. 97 is the amino acid sequence of Tb38-1 Peptide 5.
 - SEQ. ID NO. 98 is the amino acid sequence of Tb38-1 Peptide 6.
 - SEQ. ID NO. 99 is the DNA sequence of DPAS.
 - SEQ. ID NO. 100 is the deduced amino acid sequence of DPAS.
 - SEQ. ID NO. 101 is the DNA sequence of DPV.
- SEQ. ID NO. 102 is the deduced amino acid sequence of DPV.

- SEQ. ID NO. 103 is the DNA sequence of ESAT-6.
- SEQ. ID NO. 104 is the deduced amino acid sequence of ESAT-6.
- SEQ. ID NO. 105 is the DNA sequence of TbH-8-2.
- SEQ. ID NO. 106 is the DNA sequence of TbH-9FL.
- 5 SEQ. ID NO. 107 is the deduced amino acid sequence of TbH-9FL.
 - SEQ. ID NO. 108 is the DNA sequence of TbH-9-1.
 - SEQ. ID NO. 109 is the deduced amino acid sequence of TbH-9-1.
 - SEQ. ID NO. 110 is the DNA sequence of TbH-9-4.
 - SEQ. ID NO. 111 is the deduced amino acid sequence of TbH-9-4.
- SEQ. ID NO. 112 is the DNA sequence of Tb38-1F2 IN.
 - SEQ. ID NO. 113 is the DNA sequence of Tb38-2F2 RP.
 - SEQ. ID NO. 114 is the deduced amino acid sequence of Tb37-FL.
 - SEQ. ID NO. 115 is the deduced amino acid sequence of Tb38-IN.
 - SEQ. ID NO. 116 is the DNA sequence of Tb38-1F3.
- SEQ. ID NO. 117 is the deduced amino acid sequence of Tb38-1F3.
 - SEQ. ID NO. 118 is the DNA sequence of Tb38-1F5.
 - SEQ. ID NO. 119 is the DNA sequence of Tb38-1F6.
 - SEQ. ID NO. 120 is the deduced N-terminal amino acid sequence of DPV.
 - SEQ. ID NO. 121 is the deduced N-terminal amino acid sequence of AVGS.
- SEQ. ID NO. 122 is the deduced N-terminal amino acid sequence of AAMK.
 - SEQ. ID NO. 123 is the deduced N-terminal amino acid sequence of YYWC.
 - SEQ. ID NO. 124 is the deduced N-terminal amino acid sequence of DIGS.
 - SEQ. ID NO. 125 is the deduced N-terminal amino acid sequence of AEES.
 - SEQ. ID NO. 126 is the deduced N-terminal amino acid sequence of DPEP.
- SEQ. ID NO. 127 is the deduced N-terminal amino acid sequence of APKT.
 - SEQ. ID NO. 128 is the deduced amino acid sequence of DPAS.
 - SEQ. ID NO. 129 is the protein sequence of DPPD N-terminal Antigen.
 - SEQ ID NO. 130-133 are the protein sequences of four DPPD cyanogen bromide fragments.
- 30 SEQ ID NO. 134 is the N-terminal protein sequence of XDS antigen.

SEQ ID NO. 135 is the N-terminal protein sequence of AGD antigen.

SEQ ID NO. 136 is the N-terminal protein sequence of APE antigen.

SEQ ID NO. 137 is the N-terminal protein sequence of XYI antigen.

SEQ ID NO. 138 is the DNA sequence of TbH-29.

5 SEQ ID NO. 139 is the DNA sequence of TbH-30.

SEQ ID NO. 140 is the DNA sequence of TbH-32.

SEQ ID NO. 141 is the DNA sequence of TbH-33.

SEQ ID NO. 142 is the predicted amino acid sequence of TbH-29.

SEQ ID NO. 143 is the predicted amino acid sequence of TbH-30.

SEQ ID NO. 144 is the predicted amino acid sequence of TbH-32.

SEQ ID NO. 145 is the predicted amino acid sequence of TbH-33.

SEQ ID NO: 146-151 are PCR primers used in the preparation of a fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO: 152 is the DNA sequence of the fusion protein containing TbRa3,

15 38 kD and Tb38-1.

SEQ ID NO: 153 is the amino acid sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO: 154 is the DNA sequence of the M. tuberculosis antigen 38 kD.

SEQ ID NO: 155 is the amino acid sequence of the M. tuberculosis antigen 38

20 kD.

SEQ ID NO: 156 is the DNA sequence of XP14.

SEQ ID NO: 157 is the DNA sequence of XP24.

SEQ ID NO: 158 is the DNA sequence of XP31.

SEQ ID NO: 159 is the 5' DNA sequence of XP32.

SEQ ID NO: 160 is the 3' DNA sequence of XP32.

SEQ ID NO: 161 is the predicted amino acid sequence of XP14.

SEQ ID NO: 162 is the predicted amino acid sequence encoded by the reverse complement of XP14.

SEQ ID NO: 163 is the DNA sequence of XP27.

SEQ ID NO: 164 is the DNA sequence of XP36.

SEQ ID NO: 165 is the 5' DNA sequence of XP4.

SEQ ID NO: 166 is the 5' DNA sequence of XP5.

SEQ ID NO: 167 is the 5' DNA sequence of XP17.

SEQ ID NO: 168 is the 5' DNA sequence of XP30.

5 SEQ ID NO: 169 is the 5' DNA sequence of XP2.

SEQ ID NO: 170 is the 3' DNA sequence of XP2.

SEQ ID NO: 171 is the 5' DNA sequence of XP3.

SEQ ID NO: 172 is the 3' DNA sequence of XP3.

SEQ ID NO: 173 is the 5' DNA sequence of XP6.

SEQ ID NO: 174 is the 3' DNA sequence of XP6.

SEQ ID NO: 175 is the 5' DNA sequence of XP18.

SEQ ID NO: 176 is the 3' DNA sequence of XP18.

SEQ ID NO: 177 is the 5' DNA sequence of XP19.

SEQ ID NO: 178 is the 3' DNA sequence of XP19.

SEQ ID NO: 179 is the 5' DNA sequence of XP22.

SEQ ID NO: 180 is the 3' DNA sequence of XP22.

SEQ ID NO: 181 is the 5° DNA sequence of XP25.

SEQ ID NO: 182 is the 3' DNA sequence of XP25.

SEQ ID NO: 183 is the full-length DNA sequence of TbH4-XP1.

SEQ ID NO: 184 is the predicted amino acid sequence of TbH4-XP1.

SEQ ID NO: 185 is the predicted amino acid sequence encoded by the reverse complement of TbH4-XP1

SEQ ID NO: 186 is a first predicted amino acid sequence encoded by XP36.

SEQ ID NO: 187 is a second predicted amino acid sequence encoded by XP36.

SEQ ID NO: 188 is the predicted amino acid sequence encoded by XP36. complement of XP36.

SEQ ID NO: 189 is the DNA sequence of RDIF2.

SEQ ID NO: 190 is the DNA sequence of RDIF5.

SEQ ID NO: 191 is the DNA sequence of RDIF8.

SEQ ID NO: 192 is the DNA sequence of RDIF10.

SEQ ID NO: 193 is the DNA sequence of RDIF11.

SEQ ID NO: 194 is the predicted amino acid sequence of RDIF2.

SEQ ID NO: 195 is the predicted amino acid sequence of RDIF5.

SEQ ID NO: 196 is the predicted amino acid sequence of RDIF8.

SEQ ID NO: 197 is the predicted amino acid sequence of RDIF10.

SEQ ID NO: 198 is the predicted amino acid sequence of RDIF11.

SEQ ID NO: 199 is the 5' DNA sequence of RDIF12.

SEQ ID NO: 200 is the 3' DNA sequence of RDIF12.

SEQ ID NO: 201 is the DNA sequence of RDIF7.

SEQ ID NO: 202 is the predicted amino acid sequence of RDIF7.

SEQ ID NO: 203 is the DNA sequence of DIF2-1.

SEQ ID NO: 204 is the predicted amino acid sequence of DIF2-1.

SEQ ID NO: 205-212 are PCR primers used in the preparation of a fusion protein containing TbRa3, 38 kD, Tb38-1 and DPEP (hereinafter referred to as

15 TbF-2).

SEQ ID NO: 213 is the DNA sequence of the fusion protein TbF-2.

SEQ ID NO: 214 is the amino acid sequence of the fusion protein TbF-2.

SEQ ID NO: 215 is the 5' DNA sequence of MO-1.

SEQ ID NO: 216 is the 5' DNA sequence for MO-2

SEQ ID NO: 217 is the 5' DNA sequence for MO-4.

SEQ ID NO: 218 is the 5' DNA sequence for MO-8.

SEQ ID NO: 219 is the 5' DNA sequence for MO-9.

SEQ ID NO: 220 is the 5' DNA sequence for MO-26.

SEQ ID NO: 221 is the 5' DNA sequence for MO-28.

SEQ ID NO: 222 is the 5' DNA sequence for MO-29.

SEQ ID NO: 223 is the 5' DNA sequence for MO-30.

SEQ ID NO: 224 is the 5' DNA sequence for MO-34.

SEQ ID NO: 225 is the 5° DNA sequence for MO-35.

SEQ ID NO: 226 is the predicted amino acid sequence for MO-1.

SEQ ID NO: 227 is the predicted amino acid sequence for MO-2.

	SEQ ID NO: 228 is the predicted amino acid sequence for MO-4.
	SEQ ID NO: 229 is the predicted amino acid sequence for MO-8.
	SEQ ID NO: 230 is the predicted amino acid sequence for MO-9.
	SEQ ID NO: 231 is the predicted amino acid sequence for MO-26.
5	SEQ ID NO: 232 is the predicted amino acid sequence for MO-28.
	SEQ ID NO: 233 is the predicted amino acid sequence for MO-29.
	SEQ ID NO: 234 is the predicted amino acid sequence for MO-30.
	SEQ ID NO: 235 is the predicted amino acid sequence for MO-34.
	SEQ ID NO: 236 is the predicted amino acid sequence for MO-35.
10	SEQ ID NO: 237 is the determined DNA sequence for MO-10.
	SEQ ID NO: 238 is the predicted amino acid sequence for MO-10.
	SEQ ID NO: 239 is the 3' DNA sequence for MO-27.
	SEQ ID NO: 240 is the full-length DNA sequence for DPPD.
	SEQ ID NO: 241 is the predicted full-length amino acid sequence for DPPD.
15	SEQ ID NO: 242 is the determined 5' cDNA sequence for LSER-10
	SEQ ID NO: 243 is the determined 5' cDNA sequence for LSER-11
	SEQ ID NO: 244 is the determined 5' cDNA sequence for LSER-12
	SEQ ID NO: 245 is the determined 5' cDNA sequence for LSER-13
	SEQ ID NO: 246 is the determined 5' cDNA sequence for LSER-16
20	SEQ ID NO: 247 is the determined 5° cDNA sequence for LSER-25
	SEQ ID NO: 248 is the predicted amino acid sequence for LSER-10
	one of the predicted amino acid sequence for I SER 12
	SEQ ID NO: 250 is the predicted amino acid sequence for LSER-13
	SEQ 1D NO: 251 is the predicted amino acid sequence for LSEP 16
25	SEQ ID NO: 252 is the predicted amino acid sequence for LSER-25
	SEQ ID NO: 253 is the determined cDNA sequence for LSER-18
	obey 1D 190: 234 is the determined cDNA sequence for LSED 22
	SEQ 1D NO: 255 is the determined cDNA sequence for LSEP 24
	SEQ ID NO: 236 is the determined cDNA sequence for LSEP, 27
30	SEQ ID NO: 257 is the predicted amino acid sequence for LSER-18
	1 101 LOUIN-10

SEQ ID NO: 258 is the predicted amino acid sequence for LSER-23 SEQ ID NO: 259 is the predicted amino acid sequence for LSER-24 SEQ ID NO: 260 is the predicted amino acid sequence for LSER-27 SEQ ID NO: 261 is the determined 5' cDNA sequence for LSER-1 5 SEQ ID NO: 262 is the determined 5' cDNA sequence for LSER-3 SEQ ID NO: 263 is the determined 5' cDNA sequence for LSER-4 SEQ ID NO: 264 is the determined 5' cDNA sequence for LSER-5 SEQ ID NO: 265 is the determined 5' cDNA sequence for LSER-6 SEQ ID NO: 266 is the determined 5' cDNA sequence for LSER-8 10 SEQ ID NO: 267 is the determined 5° cDNA sequence for LSER-14 SEQ ID NO: 268 is the determined 5° cDNA sequence for LSER-15 SEQ ID NO: 269 is the determined 5' cDNA sequence for LSER-17 SEQ ID NO: 270 is the determined 5' cDNA sequence for LSER-19 SEQ ID NO: 271 is the determined 5' cDNA sequence for LSER-20 15 SEQ ID NO: 272 is the determined 5' cDNA sequence for LSER-22 SEQ ID NO: 273 is the determined 5' cDNA sequence for LSER-26 SEQ ID NO: 274 is the determined 5° cDNA sequence for LSER-28 SEQ ID NO: 275 is the determined 5" cDNA sequence for LSER-29 SEQ ID NO: 276 is the determined 5° cDNA sequence for LSER-30 20 SEQ ID NO: 277 is the predicted amino acid sequence for LSER-1 SEQ ID NO: 278 is the predicted amino acid sequence for LSER-3 SEQ ID NO: 279 is the predicted amino acid sequence for LSER-5 SEQ ID NO: 280 is the predicted amino acid sequence for LSER-6 SEQ ID NO: 281 is the predicted amino acid sequence for LSER-8 25 SEQ ID NO: 282 is the predicted amino acid sequence for LSER-14 SEQ ID NO: 283 is the predicted amino acid sequence for LSER-15 SEQ ID NO: 284 is the predicted amino acid sequence for LSER-17 SEQ ID NO: 285 is the predicted amino acid sequence for LSER-19 SEQ ID NO: 286 is the predicted amino acid sequence for LSER-20 30 SEQ ID NO: 287 is the predicted amino acid sequence for LSER-22

	SEQ ID NO: 288 is the predicted amino acid sequence for LSER-26
	SEQ ID NO: 289 is the predicted amino acid sequence for LSER-28
	SEQ ID NO: 290 is the predicted amino acid sequence for LSER-29
	SEQ ID NO: 291 is the predicted amino acid sequence for LSER-30
5	SEQ ID NO: 292 is the determined cDNA sequence for LSER-9
	SEQ ID NO: 293 is the determined cDNA sequence for the reverse complement
	of LSER-6
	SEQ ID NO: 294 is the predicted amino acid sequence for the reverse
	complement of LSER-6
10	SEQ ID NO: 295 is the determined 5' cDNA sequence for MO-12
	SEQ ID NO: 296 is the determined 5' cDNA sequence for MO-13
	SEQ ID NO: 297 is the determined 5' cDNA sequence for MO-19
	SEQ ID NO: 298 is the determined 5' cDNA sequence for MO-39
	SEQ ID NO: 299 is the predicted amino acid sequence for MO-12
15	SEQ ID NO: 300 is the predicted amino acid sequence for MO-13
	SEQ ID NO: 301 is the predicted amino acid sequence for MO-19
	SEQ ID NO: 302 is the predicted amino acid sequence for MO-39
	SEQ ID NO: 303 is the determined 5° cDNA sequence for Erdsn-1
	SEQ ID NO: 304 is the determined 5° cDNA sequence for Erdsn-2
20	SEQ ID NO: 305 is the determined 5' cDNA sequence for Erdsn-4
	SEQ ID NO: 306 is the determined 5' cDNA sequence for Erdsn-5
	SEQ ID NO: 307 is the determined 5' cDNA sequence for Erdsn-6
	SEQ ID NO: 308 is the determined 5' cDNA sequence for Erdsn-7
	SEQ ID NO: 309 is the determined 5' cDNA sequence for Erdsn-8
25	SEQ ID NO: 310 is the determined 5' cDNA sequence for Erdsn-9
	SEQ ID NO: 311 is the determined 5' cDNA sequence for Erdsn-10
	SEQ ID NO: 312 is the determined 5' cDNA sequence for Erdsn-12
	SEQ ID NO: 313 is the determined 5' cDNA sequence for Erdsn-13
	SEQ ID NO: 314 is the determined 5' cDNA sequence for Erdsn-14
30	SEQ ID NO: 315 is the determined 5' cDNA sequence for Frdsn-15

SEQ ID NO: 316 is the determined 5' cDNA sequence for Erdsn-16 SEQ ID NO: 317 is the determined 5' cDNA sequence for Erdsn-17 SEQ ID NO: 318 is the determined 5' cDNA sequence for Erdsn-18 SEQ ID NO: 319 is the determined 5° cDNA sequence for Erdsn-21 SEQ ID NO: 320 is the determined 5' cDNA sequence for Erdsn-22 5 SEQ ID NO: 321 is the determined 5' cDNA sequence for Erdsn-23 SEQ ID NO: 322 is the determined 5' cDNA sequence for Erdsn-25 SEQ ID NO: 323 is the determined 3' cDNA sequence for Erdsn-1 SEQ ID NO: 324 is the determined 3' cDNA sequence for Erdsn-2 SEQ ID NO: 325 is the determined 3' cDNA sequence for Erdsn-4 10 SEQ ID NO: 326 is the determined 3' cDNA sequence for Erdsn-5 SEQ ID NO: 327 is the determined 3' cDNA sequence for Erdsn-7 SEQ ID NO: 328 is the determined 3' cDNA sequence for Erdsn-8 SEQ ID NO: 329 is the determined 3' cDNA sequence for Erdsn-9 SEQ ID NO: 330 is the determined 3' cDNA sequence for Erdsn-10 15 SEQ ID NO: 331 is the determined 3' cDNA sequence for Erdsn-12 SEQ ID NO: 332 is the determined 3' cDNA sequence for Erdsn-13 SEQ ID NO: 333 is the determined 3' cDNA sequence for Erdsn-14 SEQ ID NO: 334 is the determined 3' cDNA sequence for Erdsn-15 SEQ ID NO: 335 is the determined 3' cDNA sequence for Erdsn-16 20 SEQ ID NO: 336 is the determined 3' cDNA sequence for Erdsn-17 SEQ ID NO: 337 is the determined 3' cDNA sequence for Erdsn-18 SEQ ID NO: 338 is the determined 3' cDNA sequence for Erdsn-21 SEQ ID NO: 339 is the determined 3' cDNA sequence for Erdsn-22 SEQ ID NO: 340 is the determined 3' cDNA sequence for Erdsn-23 25 SEQ ID NO: 341 is the determined 3' cDNA sequence for Erdsn-25 SEQ ID NO: 342 is the determined cDNA sequence for Erdsn-24 SEQ ID NO: 343 is the determined amino acid sequence for a M. tuberculosis 85b precursor homolog SEQ ID NO: 344 is the determined amino acid sequence for spot 1 30

25

SEQ ID NO: 345 is a determined amino acid sequence for spot 2

SEQ ID NO: 346 is a determined amino acid sequence for spot 2

SEQ ID NO: 347 is the determined amino acid seq for spot 4

SEQ ID NO: 348 is the sequence of primer PDM-157

SEQ ID NO: 349 is the sequence of primer PDM-160

SEQ ID NO: 350 is the DNA sequence of the fusion protein TbF-6

SEQ ID NO: 351 is the amino acid sequence of fusion protein TbF-6

SEQ ID NO: 352 is the sequence of primer PDM-176

SEQ ID NO: 353 is the sequence of primer PDM-175

SEQ ID NO: 354 is the DNA sequence of the fusion protein TbF-8

SEQ ID NO: 355 is the amino acid sequence of the fusion protein TbF-8

DETAILED DESCRIPTION OF THE INVENTION

As noted above, the present invention is generally directed to compositions and methods for preventing, treating and diagnosing tuberculosis. The 15 compositions of the subject invention include polypeptides that comprise at least one immunogenic portion of a M. tuberculosis antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications. Polypeptides within the scope of the present invention include, but are not limited to, immunogenic soluble M. tuberculosis antigens. A "soluble M tuberculosis antigen" is a protein of M. tuberculosis origin that is present in M. tuberculosis culture filtrate. As used herein, the term "polypeptide" encompasses amino acid chains of any length, including full length proteins (i.e., antigens), wherein the amino acid residues are linked by covalent peptide bonds. Thus, a polypeptide comprising an immunogenic portion of one of the above antigens may consist entirely of the immunogenic portion, or may contain additional sequences. The additional sequences may be derived from the native M. tuberculosis antigen or may be heterologous, and such sequences may (but need not) be immunogenic.

"Immunogenic." as used herein, refers to the ability to elicit an immune response (e.g., cellular) in a patient, such as a human, and/or in a biological sample. In particular, antigens that are immunogenic (and immunogenic portions or other variants

of such antigens) are capable of stimulating cell proliferation, interleukin-12 production and/or interferon-y production in biological samples comprising one or more cells selected from the group of T cells, NK cells, B cells and macrophages, where the cells are derived from an *M. tuberculosis*-immune individual. Polypeptides comprising at least an immunogenic portion of one or more *M. tuberculosis* antigens may generally be used to detect tuberculosis or to induce protective immunity against tuberculosis in a patient.

The compositions and methods of the present invention also encompass variants of the above polypeptides and DNA molecules. A polypeptide "variant," as used herein, is a polypeptide that differs from the recited polypeptide only in conservative substitutions and/or modifications, such that the therapeutic, antigenic and/or immunogenic properties of the polypeptide are retained. Polypeptide variants preferably exhibit at least about 70%, more preferably at least about 90% and most preferably at least about 95% identity to the identified polypeptides. For polypeptides with immunoreactive properties, variants may, alternatively, be identified by modifying the amino acid sequence of one of the above polypeptides, and evaluating the immunoreactivity of the modified polypeptide. For polypeptides useful for the generation of diagnostic binding agents, a variant may be identified by evaluating a modified polypeptide for the ability to generate antibodies that detect the presence or absence of tuberculosis. Such modified sequences may be prepared and tested using, for example, the representative procedures described herein.

10

20

As used herein, a "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. In general, the following groups of amino acids represent conservative changes: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his.

Variants may also, or alternatively, contain other modifications, including the deletion or addition of amino acids that have minimal influence on the

21

antigenic properties, secondary structure and hydropathic nature of the polypeptide. For example, a polypeptide may be conjugated to a signal (or leader) sequence at the N-terminal end of the protein which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (e.g., poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

A nucleotide "variant" is a sequence that differs from the recited nucleotide sequence in having one or more nucleotide deletions, substitutions or additions. Such modifications may be readily introduced using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis as taught, for example, by Adelman et al. (DNA, 2:183, 1983). Nucleotide variants may be naturally occurring allelic variants, or non-naturally occurring variants. Variant nucleotide sequences preferably exhibit at least about 70%, more preferably at least about 80% and most preferably at least about 90% identity to the recited sequence. Such variant nucleotide sequences will generally hybridize to the recite nucleotide sequence under stringent conditions. As used herein, "stringent conditions" refers to prewashing in a solution of 6X SSC, 0.2% SDS; hybridizing at 65 °C, 6X SSC, 0.2% SDS overnight: followed by two washes of 30 minutes each in 1X SSC, 0.1% SDS at 65 °C and two washes of 30 minutes each in 0.2X SSC, 0.1% SDS at 65 °C.

10

20

30

In a related aspect, combination polypeptides are disclosed. A "combination polypeptide" is a polypeptide comprising at least one of the above immunogenic portions and one or more additional immunogenic *M. tuberculosis* sequences, which are joined via a peptide linkage into a single amino acid chain. The sequences may be joined directly (*i.e.*, with no intervening amino acids) or may be joined by way of a linker sequence (*e.g.*, Gly-Cys-Gly) that does not significantly diminish the immunogenic properties of the component polypeptides.

In general, *M. tuberculosis* antigens, and DNA sequences encoding such antigens, may be prepared using any of a variety of procedures. For example, soluble antigens may be isolated from *M. tuberculosis* culture filtrate by procedures known to

22

those of ordinary skill in the art, including anion-exchange and reverse phase chromatography. Purified antigens are then evaluated for their ability to elicit an appropriate immune response (e.g., cellular) using, for example, the representative methods described herein. Immunogenic antigens may then be partially sequenced using techniques such as traditional Edman chemistry. See Edman and Berg, Eur. J. Biochem. 80:116-132, 1967.

Immunogenic antigens may also be produced recombinantly using a DNA sequence that encodes the antigen, which has been inserted into an expression vector and expressed in an appropriate host. DNA molecules encoding soluble antigens may be isolated by screening an appropriate *M. tuberculosis* expression library with anti-sera (e.g., rabbit) raised specifically against soluble *M. tuberculosis* antigens. DNA sequences encoding antigens that may or may not be soluble may be identified by screening an appropriate *M. tuberculosis* genomic or cDNA expression library with sera obtained from patients infected with *M. tuberculosis*. Such screens may generally be performed using techniques well known to those of ordinary skill in the art, such as those described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories. Cold Spring Harbor, NY, 1989.

DNA sequences encoding soluble antigens may also be obtained by screening an appropriate *M. tuberculosis* cDNA or genomic DNA library for DNA sequences that hybridize to degenerate oligonucleotides derived from partial amino acid sequences of isolated soluble antigens. Degenerate oligonucleotide sequences for use in such a screen may be designed and synthesized, and the screen may be performed, as described (for example) in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989 (and references cited therein). Polymerase chain reaction (PCR) may also be employed, using the above oligonucleotides in methods well known in the art, to isolate a nucleic acid probe from a cDNA or genomic library. The library screen may then be performed using the isolated probe.

20

25

Alternatively, genomic or cDNA libraries derived from M. tuberculosis
may be screened directly using peripheral blood mononuclear cells (PBMCs) or T cell

23

lines or clones derived from one or more *M. tuberculosis*-immune individuals. In general, PBMCs and/or T cells for use in such screens may be prepared as described below. Direct library screens may generally be performed by assaying pools of expressed recombinant proteins for the ability to induce proliferation and/or interferon- γ production in T cells derived from an *M. tuberculosis*-immune individual. Alternatively, potential T cell antigens may be first selected based on antibody reactivity, as described above.

Regardless of the method of preparation, the antigens (and immunogenic portions thereof) described herein (which may or may not be soluble) have the ability to induce an immunogenic response. More specifically, the antigens have the ability to induce proliferation and/or cytokine production (i.e., interferon-y and/or interleukin-12 production) in T cells, NK cells, B cells and/or macrophages derived from an M. tuberculosis-immune individual. The selection of cell type for use in evaluating an immunogenic response to a antigen will, of course, depend on the desired response. For example, interleukin-12 production is most readily evaluated using preparations containing B cells and/or macrophages. An M. tuberculosis-immune individual is one who is considered to be resistant to the development of tuberculosis by virtue of having mounted an effective T cell response to M. tuberculosis (i.e., substantially free of disease symptoms). Such individuals may be identified based on a strongly positive (i.e., greater than about 10 mm diameter induration) intradermal skin test response to tuberculosis proteins (PPD) and an absence of any signs or symptoms of tuberculosis disease. T cells, NK cells, B cells and macrophages derived from M. tuberculosisimmune individuals may be prepared using methods known to those of ordinary skill in the art. For example, a preparation of PBMCs (i.e., peripheral blood mononuclear cells) may be employed without further separation of component cells. PBMCs may generally be prepared, for example, using density centrifugation through FicollTM (Winthrop Laboratories, NY). T cells for use in the assays described herein may also be purified directly from PBMCs. Alternatively, an enriched T cell line reactive against mycobacterial proteins, or T cell clones reactive to individual mycobacterial proteins. may be employed. Such T cell clones may be generated by, for example, culturing

20

30

24

PBMCs from *M. tuberculosis*-immune individuals with mycobacterial proteins for a period of 2-4 weeks. This allows expansion of only the mycobacterial protein-specific T cells, resulting in a line composed solely of such cells. These cells may then be cloned and tested with individual proteins, using methods known to those of ordinary skill in the art, to more accurately define individual T cell specificity. In general, antigens that test positive in assays for proliferation and/or cytokine production (*i.e.*, interferon-γ and/or interleukin-12 production) performed using T cells, NK cells. B cells and/or macrophages derived from an *M. tuberculosis*-immune individual are considered immunogenic. Such assays may be performed, for example, using the representative procedures described below. Immunogenic portions of such antigens may be identified using similar assays, and may be present within the polypeptides described herein.

5

20

25

The ability of a polypeptide (e.g., an immunogenic antigen, or a portion or other variant thereof) to induce cell proliferation is evaluated by contacting the cells (e.g., T cells and/or NK cells) with the polypeptide and measuring the proliferation of the cells. In general, the amount of polypeptide that is sufficient for evaluation of about 10⁵ cells ranges from about 10 ng/mL to about 100 µg/mL and preferably is about 10 µg/mL. The incubation of polypeptide with cells is typically performed at 37°C for about six days. Following incubation with polypeptide, the cells are assayed for a proliferative response, which may be evaluated by methods known to those of ordinary skill in the art, such as exposing cells to a pulse of radiolabeled thymidine and measuring the incorporation of label into cellular DNA. In general, a polypeptide that results in at least a three fold increase in proliferation above background (i.e., the proliferation observed for cells cultured without polypeptide) is considered to be able to induce proliferation.

The ability of a polypeptide to stimulate the production of interferon- γ and/or interleukin-12 in cells may be evaluated by contacting the cells with the polypeptide and measuring the level of interferon- γ or interleukin-12 produced by the cells. In general, the amount of polypeptide that is sufficient for the evaluation of about 10^5 cells ranges from about 10 ng/mL to about 100 µg/mL and preferably is about 10 µg/mL. The polypeptide may, but need not, be immobilized on a solid support, such as a

bead or a biodegradable microsphere, such as those described in U.S. Patent Nos. 4,897,268 and 5,075,109. The incubation of polypeptide with the cells is typically performed at 37°C for about six days. Following incubation with polypeptide, the cells are assayed for interferon-γ and/or interleukin-12 (or one or more subunits thereof), which may be evaluated by methods known to those of ordinary skill in the art, such as an enzyme-linked immunosorbent assay (ELISA) or, in the case of IL-12 P70 subunit, a bioassay such as an assay measuring proliferation of T cells. In general, a polypeptide that results in the production of at least 50 pg of interferon-γ per mL of cultured supernatant (containing 10⁴-10⁵ T cells per mL) is considered able to stimulate the production of interferon-γ. A polypeptide that stimulates the production of at least 10 pg/mL of IL-12 P70 subunit, and/or at least 100 pg/mL of IL-12 P40 subunit, per 10³ macrophages or B cells (or per 3 x 10⁵ PBMC) is considered able to stimulate the production of IL-12.

In general, immunogenic antigens are those antigens that stimulate proliferation and/or cytokine production (i.e., interferon-y and/or interleukin-12 production) in T cells, NK cells, B cells and/or macrophages derived from at least about 25% of M. tuberculosis-immune individuals. Among these immunogenic antigens, polypeptides having superior therapeutic properties may be distinguished based on the magnitude of the responses in the above assays and based on the percentage of individuals for which a response is observed. In addition, antigens having superior therapeutic properties will not stimulate proliferation and/or cytokine production in vitro in cells derived from more than about 25% of individuals that are not M. tuberculosis-immune, thereby eliminating responses that are not specifically due to M. tuberculosis-responsive cells. Those antigens that induce a response in a high percentage of T cell, NK cell. B cell and/or macrophage preparations from M. tuberculosis-immune individuals (with a low incidence of responses in cell preparations from other individuals) have superior therapeutic properties.

20

30

Antigens with superior therapeutic properties may also be identified based on their ability to diminish the severity of *M. tuberculosis* infection in experimental animals, when administered as a vaccine. Suitable vaccine preparations

for use on experimental animals are described in detail below. Efficacy may be determined based on the ability of the antigen to provide at least about a 50% reduction in bacterial numbers and/or at least about a 40% decrease in mortality following experimental infection. Suitable experimental animals include mice, guinea pigs and primates.

Antigens having superior diagnostic properties may generally be identified based on the ability to elicit a response in an intradermal skin test performed on an individual with active tuberculosis, but not in a test performed on an individual who is not infected with *M. tuberculosis*. Skin tests may generally be performed as described below, with a response of at least 5 mm induration considered positive.

Immunogenic portions of the antigens described herein may be prepared and identified using well known techniques, such as those summarized in Paul, Fundamental Immunology, 3d ed., Raven Press, 1993, pp. 243-247 and references cited therein. Such techniques include screening polypeptide portions of the native antigen for immunogenic properties. The representative proliferation and cytokine production assays described herein may generally be employed in these screens. An immunogenic portion of a polypeptide is a portion that, within such representative assays, generates an immune response (e.g., proliferation, interferon-y production and/or interleukin-12 production) that is substantially similar to that generated by the full length antigen. In other words, an immunogenic portion of an antigen may generate at least about 20%, and preferably about 100%, of the proliferation induced by the full length antigen in the model proliferation assay described herein. An immunogenic portion may also, or alternatively, stimulate the production of at least about 20%, and preferably about 100%, of the interferon-y and/or interleukin-12 induced by the full length antigen in the model assay described herein.

15

20

30

Portions and other variants of *M. tuberculosis* antigens may be generated by synthetic or recombinant means. Synthetic polypeptides having fewer than about 100 amino acids, and generally fewer than about 50 amino acids, may be generated using techniques well known to those of ordinary skill in the art. For example, such polypeptides may be synthesized using any of the commercially available solid-phase

techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. See Merrifield, J. Am. Chem. Soc. 85:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Applied BioSystems, Inc., Foster City, CA, and may be operated according to the manufacturer's instructions. Variants of a native antigen may generally be prepared using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis. Sections of the DNA sequence may also be removed using standard techniques to permit preparation of truncated polypeptides.

Recombinant polypeptides containing portions and/or variants of a native antigen may be readily prepared from a DNA sequence encoding the polypeptide using a variety of techniques well known to those of ordinary skill in the art. For example, supernatants from suitable host/vector systems which secrete recombinant protein into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant protein.

Any of a variety of expression vectors known to those of ordinary skill in the art may be employed to express recombinant polypeptides of this invention. Expression may be achieved in any appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast and higher eukaryotic cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line such as COS or CHO. The DNA sequences expressed in this manner may encode naturally occurring antigens, portions of naturally occurring antigens, or other variants thereof.

In general, regardless of the method of preparation, the polypeptides disclosed herein are prepared in substantially pure form. Preferably, the polypeptides are at least about 80% pure, more preferably at least about 90% pure and most

preferably at least about 99% pure. In certain preferred embodiments, described in detail below, the substantially pure polypeptides are incorporated into pharmaceutical compositions or vaccines for use in one or more of the methods disclosed herein.

In certain specific embodiments, the subject invention discloses polypeptides comprising at least an immunogenic portion of a soluble *M. tuberculosis* antigen having one of the following N-terminal sequences, or a variant thereof that differs only in conservative substitutions and/or modifications:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 120)
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser; (SEQ ID No. 121)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 122)
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 123)
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val; (SEQ ID No. 124)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID No. 125)
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Ala-Ala-Ala-Ser-Pro-Pro-Ser; (SEQ ID No. 126)
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 127)
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn; (SEQ ID No. 128)
- (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 134)
- (k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID No. 135) or

10

15

20

25

30

25

30

(l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136)

wherein Xaa may be any amino acid, preferably a cysteine residue. A DNA sequence encoding the antigen identified as (g) above is provided in SEQ ID No. 52, and the polypeptide encoded by SEQ ID No. 52 is provided in SEQ ID No. 53. A DNA sequence encoding the antigen defined as (a) above is provided in SEQ ID No. 101; its deduced amino acid sequence is provided in SEQ ID No. 102. A DNA sequence corresponding to antigen (d) above is provided in SEQ ID No. 24 a DNA sequence corresponding to antigen (c) is provided in SEQ ID No. 25 and a DNA sequence corresponding to antigen (i) is provided in SEQ ID No. 99; its deduced amino acid sequence is provided in SEQ ID No. 99; its deduced amino acid

In a further specific embodiment, the subject invention discloses polypeptides comprising at least an immunogenic portion of an *M. tuberculosis* antigen having one of the following N-terminal sequences, or a variant thereof that differs only in conservative substitutions and/or modifications:

- (m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No 137) or
- (n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID No. 129)
- wherein Xaa may be any amino acid, preferably a cysteine residue.

In other specific embodiments, the subject invention discloses polypeptides comprising at least an immunogenic portion of a soluble *M. tuberculosis* antigen (or a variant of such an antigen) that comprises one or more of the amino acid sequences encoded by (a) the DNA sequences of SEQ ID Nos.: 1, 2, 4-10, 13-25 and 52; (b) the complements of such DNA sequences, or (c) DNA sequences substantially homologous to a sequence in (a) or (b).

In further specific embodiments, the subject invention discloses polypeptides comprising at least an immunogenic portion of a *M. tuberculosis* antigen (or a variant of such an antigen), which may or may not be soluble, that comprises one or more of the amino acid sequences encoded by (a) the DNA sequences of SEQ ID

10

25

Nos.: 26-51, 138, 139, 163-183, 189-193, 199, 200, 201, 203, 215-225, 239, 240, 242-247, 253-256, 261-276, 292, 293, 295-298 and 303-342, (b) the complements of such DNA sequences or (c) DNA sequences substantially homologous to a sequence in (a) or (b).

In the specific embodiments discussed above, the *M. tuberculosis* antigens include variants that are encoded by DNA sequences which are substantially homologous to one or more of DNA sequences specifically recited herein. "Substantial homology," as used herein, refers to DNA sequences that are capable of hybridizing under moderately stringent conditions. Suitable moderately stringent conditions include prewashing in a solution of 5X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-65°C, 5X SSC, overnight or, in the case of cross-species homology at 45°C, 0.5X SSC; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS). Such hybridizing DNA sequences are also within the scope of this invention, as are nucleotide sequences that, due to code degeneracy, encode an immunogenic polypeptide that is encoded by a hybridizing DNA sequence.

In a related aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, a polypeptide of the present invention and a known *M tuberculosis* antigen, such as the 38 kD antigen described in Andersen and Hansen, *Infect. Immun.* 57:2481-2488, 1989, (Genbank Accession No. M30046) or ESAT-6 (SEQ ID Nos. 103 and 104), together with variants of such fusion proteins. The fusion proteins of the present invention may also include a linker peptide between the first and second polypeptides.

A DNA sequence encoding a fusion protein of the present invention is constructed using known recombinant DNA techniques to assemble separate DNA sequences encoding the first and second polypeptides into an appropriate expression vector. The 3' end of a DNA sequence encoding the first polypeptide is ligated, with or without a peptide linker, to the 5'-end of a DNA sequence encoding the second polypeptide so that the reading frames of the sequences are in phase to permit mRNA

31

translation of the two DNA sequences into a single fusion protein that retains the biological activity of both the first and the second polypeptides.

A peptide linker sequence may be employed to separate the first and the second polypeptides by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into 5 the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., Gene 40:39-46, 1985; Murphy et al., Proc. Natl. Acad. Sci. USA 83:8258-8262, 1986; U.S. Patent No. 4,935.233 and U.S. Patent No. 4,751,180. The linker sequence may be from 1 to about 50 amino acids in length. Peptide sequences are not required when the first and second polypeptides have nonessential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons require to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

20

In another aspect, the present invention provides methods for using one or more of the above polypeptides or fusion proteins (or DNA molecules encoding such polypeptides) to induce protective immunity against tuberculosis in a patient. As used herein, a "patient" refers to any warm-blooded animal, preferably a human. A patient

32

may be afflicted with a disease, or may be free of detectable disease and/or infection. In other words, protective immunity may be induced to prevent or treat tuberculosis.

In this aspect, the polypeptide, fusion protein or DNA molecule is generally present within a pharmaceutical composition and/or a vaccine.

Pharmaceutical compositions may comprise one or more polypeptides, each of which may contain one or more of the above sequences (or variants thereof), and a physiologically acceptable carrier. Vaccines may comprise one or more of the above polypeptides and a non-specific immune response enhancer, such as an adjuvant or a liposome (into which the polypeptide is incorporated). Such pharmaceutical compositions and vaccines may also contain other *M. tuberculosis* antigens, either incorporated into a combination polypeptide or present within a separate polypeptide.

Alternatively, a vaccine may contain DNA encoding one or more polypeptides as described above, such that the polypeptide is generated in situ. In such vaccines, the DNA may be present within any of a variety of delivery systems known to those of ordinary skill in the art, including nucleic acid expression systems, bacterial and viral expression systems. Appropriate nucleic acid expression systems contain the necessary DNA sequences for expression in the patient (such as a suitable promoter and terminating signal). Bacterial delivery systems involve the administration of a bacterium (such as Bacillus-Calmette-Guerrin) that expresses an immunogenic portion of the polypeptide on its cell surface. In a preferred embodiment, the DNA may be introduced using a viral expression system (e.g., vaccinia or other pox virus, retrovirus, or adenovirus), which may involve the use of a non-pathogenic (defective), replication competent virus. Techniques for incorporating DNA into such expression systems are well known to those of ordinary skill in the art. The DNA may also be "naked." as described, for example, in Ulmer et al., Science 259:1745-1749, 1993 and reviewed by Cohen, Science 259:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

15

In a related aspect, a DNA vaccine as described above may be administered simultaneously with or sequentially to either a polypeptide of the present

5

10

15

20

33

invention or a known *M. tuberculosis* antigen, such as the 38 kD antigen described above. For example, administration of DNA encoding a polypeptide of the present invention, either "naked" or in a delivery system as described above, may be followed by administration of an antigen in order to enhance the protective immune effect of the vaccine.

Routes and frequency of administration, as well as dosage, will vary from individual to individual and may parallel those currently being used in immunization using BCG. In general, the pharmaceutical compositions and vaccines may be administered by injection (e.g., intracutaneous, intramuscular, intravenous or subcutaneous), intranasally (e.g., by aspiration) or orally. Between 1 and 3 doses may be administered for a 1-36 week period. Preferably, 3 doses are administered, at intervals of 3-4 months, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of polypeptide or DNA that, when administered as described above, is capable of raising an immune response in an immunized patient sufficient to protect the patient from M. tuberculosis infection for at least 1-2 years. In general, the amount of polypeptide present in a dose (or produced in situ by the DNA in a dose) ranges from about 1 pg to about 100 mg per kg of host, typically from about 10 pg to about 1 mg, and preferably from about 100 pg to about 1 µg. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

While any suitable carrier known to those of ordinary skill in the art may be employed in the pharmaceutical compositions of this invention, the type of carrier will vary depending on the mode of administration. For parenteral administration, such as subcutaneous injection, the carrier preferably comprises water, saline, alcohol, a fat, a wax or a buffer. For oral administration, any of the above carriers or a solid carrier, such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, talcum, cellulose, glucose, sucrose, and magnesium carbonate, may be employed. Biodegradable microspheres (e.g., polylactic galactide) may also be employed as carriers for the pharmaceutical compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268 and 5,075,109.

34

Any of a variety of adjuvants may be employed in the vaccines of this invention to nonspecifically enhance the immune response. Most adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a nonspecific stimulator of immune responses, such as lipid A, Bortadella pertussis or Mycobacterium tuberculosis. Suitable adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Freund's Complete Adjuvant (Difco Laboratories) and Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ). Other suitable adjuvants include alum, biodegradable microspheres, monophosphoryl lipid A and quil A.

In another aspect, this invention provides methods for using one or more of the polypeptides described above to diagnose tuberculosis using a skin test. As used herein, a "skin test" is any assay performed directly on a patient in which a delayed-type hypersensitivity (DTH) reaction (such as swelling, reddening or dermatitis) is measured following intradermal injection of one or more polypeptides as described above. Such injection may be achieved using any suitable device sufficient to contact the polypeptide or polypeptides with dermal cells of the patient, such as a tuberculin syringe or 1 mL syringe. Preferably, the reaction is measured at least 48 hours after injection, more preferably 48-72 hours.

10

15

20

25

30

The DTH reaction is a cell-mediated immune response, which is greater in patients that have been exposed previously to the test antigen (i.e., the immunogenic portion of the polypeptide employed, or a variant thereof). The response may be measured visually, using a ruler. In general, a response that is greater than about 0.5 cm in diameter, preferably greater than about 1.0 cm in diameter, is a positive response, indicative of tuberculosis infection, which may or may not be manifested as an active disease.

The polypeptides of this invention are preferably formulated, for use in a skin test, as pharmaceutical compositions containing a polypeptide and a physiologically acceptable carrier, as described above. Such compositions typically contain one or more of the above polypeptides in an amount ranging from about 1 µg to about 100 µg, preferably from about 10 µg to about 50 µg in a volume of 0.1 mL.

Preferably, the carrier employed in such pharmaceutical compositions is a saline solution with appropriate preservatives, such as phenol and/or Tween 80TM.

In a preferred embodiment, a polypeptide employed in a skin test is of sufficient size such that it remains at the site of injection for the duration of the reaction period. In general, a polypeptide that is at least 9 amino acids in length is sufficient. The polypeptide is also preferably broken down by macrophages within hours of injection to allow presentation to T-cells. Such polypeptides may contain repeats of one or more of the above sequences and/or other immunogenic or nonimmunogenic sequences.

10

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLES

15

EXAMPLE 1

PURIFICATION AND CHARACTERIZATION OF POLYPEPTIDES FROM M. TUBERCULOSIS CULTURE FILTRATE

This example illustrates the preparation of *M. tuberculosis* soluble polypeptides from culture filtrate. Unless otherwise noted, all percentages in the following example are weight per volume.

M. tuberculosis (either H37Ra, ATCC No. 25177, or H37Rv, ATCC No. 25618) was cultured in sterile GAS media at 37°C for fourteen days. The media was then vacuum filtered (leaving the bulk of the cells) through a 0.45 μ filter into a sterile 2.5 L bottle. The media was next filtered through a 0.2 μ filter into a sterile 4 L bottle and NaN₃ was added to the culture filtrate to a concentration of 0.04%. The bottles were then placed in a 4°C cold foom.

The culture filtrate was concentrated by placing the filtrate in a 12 L reservoir that had been autoclaved and feeding the filtrate into a 400 ml Amicon stir cell

WO 99/42076 PCT/US99/03268

which had been rinsed with ethanol and contained a 10,000 kDa MWCO membrane. The pressure was maintained at 60 psi using nitrogen gas. This procedure reduced the 12 L volume to approximately 50 ml.

The culture filtrate was dialyzed into 0.1% ammonium bicarbonate using a 8,000 kDa MWCO cellulose ester membrane, with two changes of ammonium bicarbonate solution. Protein concentration was then determined by a commercially available BCA assay (Pierce, Rockford, IL).

The dialyzed culture filtrate was then lyophilized, and the polypeptides resuspended in distilled water. The polypeptides were dialyzed against 0.01 mM 1.3 bis[tris(hydroxymethyl)-methylamino]propane. pH 7.5 (Bis-Tris propane buffer), the initial conditions for anion exchange chromatography. Fractionation was performed using gel profusion chromatography on a POROS 146 II Q/M anion exchange column 4.6 mm x 100 mm (Perseptive BioSystems, Framingham, MA) equilibrated in 0.01 mM Bis-Tris propane buffer pH 7.5. Polypeptides were eluted with a linear 0-0.5 M NaCl gradient in the above buffer system. The column eluent was monitored at a wavelength of 220 nm.

10

15

25

The pools of polypeptides eluting from the ion exchange column were dialyzed against distilled water and lyophilized. The resulting material was dissolved in 0.1% trifluoroacetic acid (TFA) pH 1.9 in water, and the polypeptides were purified on a Delta-Pak C18 column (Waters, Milford, MA) 300 Angstrom pore size, 5 micron particle size (3.9 x 150 mm). The polypeptides were eluted from the column with a linear gradient from 0-60% dilution buffer (0.1% TFA in acetonitrile). The flow rate was 0.75 ml/minute and the HPLC eluent was monitored at 214 nm. Fractions containing the eluted polypeptides were collected to maximize the purity of the individual samples. Approximately 200 purified polypeptides were obtained.

The purified polypeptides were then screened for the ability to induce T-cell proliferation in PBMC preparations. The PBMCs from donors known to be PPD skin test positive and whose T-cells were shown to proliferate in response to PPD and crude soluble proteins from MTB were cultured in medium comprising RPMI 1640 supplemented with 10% pooled human serum and $50 \,\mu\text{g/ml}$ gentamicin. Purified

polypeptides were added in duplicate at concentrations of 0.5 to $10 \,\mu\text{g/mL}$. After six days of culture in 96-well round-bottom plates in a volume of $200 \,\mu\text{l}$, $50 \,\mu\text{l}$ of medium was removed from each well for determination of IFN- γ levels, as described below. The plates were then pulsed with $1 \,\mu\text{Ci/well}$ of tritiated thymidine for a further 18 hours, harvested and tritium uptake determined using a gas scintillation counter. Fractions that resulted in proliferation in both replicates three fold greater than the proliferation observed in cells cultured in medium alone were considered positive.

IFN-y was measured using an enzyme-linked immunosorbent assay (ELISA). ELISA plates were coated with a mouse monoclonal antibody directed to human IFN-y (PharMingen, San Diego, CA) in PBS for four hours at room temperature. Wells were then blocked with PBS containing 5% (W/V) non-fat dried milk for 1 hour at room temperature. The plates were then washed six times in PBS/0.2% TWEEN-20 and samples diluted 1:2 in culture medium in the ELISA plates were incubated overnight at room temperature. The plates were again washed and a polyclonal rabbit anti-human IFN-y serum diluted 1:3000 in PBS/10% normal goat serum was added to each well. The plates were then incubated for two hours at room temperature, washed and horseradish peroxidase-coupled anti-rabbit IgG (Sigma Chemical So., St. Louis. MO) was added at a 1:2000 dilution in PBS/5% non-fat dried milk. After a further two hour incubation at room temperature, the plates were washed and TMB substrate added. The reaction was stopped after 20 min with 1 N sulfuric acid. Optical density was determined at 450 nm using 570 nm as a reference wavelength. Fractions that resulted in both replicates giving an OD two fold greater than the mean OD from cells cultured in medium alone, plus 3 standard deviations, were considered positive.

10

15

For sequencing, the polypeptides were individually dried onto Biobrene

TM (Perkin Elmer/Applied BioSystems Division, Foster City, CA) treated glass fiber filters. The filters with polypeptide were loaded onto a Perkin Elmer/Applied BioSystems Division Procise 492 protein sequencer. The polypeptides were sequenced from the amino terminal and using traditional Edman chemistry. The amino acid sequence was determined for each polypeptide by comparing the retention time of the PTH amino acid derivative to the appropriate PTH derivative standards.

Using the procedure described above, antigens having the following N-terminal sequences were isolated:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Xaa-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 54)
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser; (SEQ ID No. 55)
 - (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 56)
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 57)
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val; (SEQ ID No. 58)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID No. 59)
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Ala-Ala-Ala-Ala-Pro-Pro-Ala; (SEQ ID No. 60) and
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 61)

wherein Xaa may be any amino acid.

5

10

15

An additional antigen was isolated employing a microbore HPLC purification step in addition to the procedure described above. Specifically, 20 µl of a fraction comprising a mixture of antigens from the chromatographic purification step previously described, was purified on an Aquapore C18 column (Perkin Elmer/Applied Biosystems Division, Foster City, CA) with a 7 micron pore size, column size 1 mm x 100 mm, in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted from the column with a linear gradient of 1%/minute of acetonitrile (containing 0.05% TFA) in water (0.05% TFA) at a flow rate of 80 µl/minute. The eluent was monitored at 250 nm. The original fraction was separated into 4 major peaks plus other smaller components and a polypeptide was obtained which was shown to

10

15

20

25

30

have a molecular weight of 12.054 Kd (by mass spectrometry) and the following N-terminal sequence:

 (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Gln-Thr-Ser-Leu-Leu-Asn-Asn-Leu-Ala-Asp-Pro-Asp-Val-Ser-Phe-Ala-Asp (SEQ ID No. 62).

This polypeptide was shown to induce proliferation and IFN-y production in PBMC preparations using the assays described above.

Additional soluble antigens were isolated from *M. tuberculosis* culture filtrate as follows. *M. tuberculosis* culture filtrate was prepared as described above. Following dialysis against Bis-Tris propane buffer, at pH 5.5, fractionation was performed using anion exchange chromatography on a Poros QE column 4.6 x 100 mm (Perseptive Biosystems) equilibrated in Bis-Tris propane buffer pH 5.5. Polypeptides were eluted with a linear 0-1.5 M NaCl gradient in the above buffer system at a flow rate of 10 ml/min. The column eluent was monitored at a wavelength of 214 nm.

The fractions eluting from the ion exchange column were pooled and subjected to reverse phase chromatography using a Poros R2 column 4.6 x 100 mm (Perseptive Biosystems). Polypeptides were eluted from the column with a linear gradient from 0-100% acetonitrile (0.1% TFA) at a flow rate of 5 ml/min. The eluent was monitored at 214 nm.

Fractions containing the eluted polypeptides were lyophilized and resuspended in 80 µl of aqueous 0.1% TFA and further subjected to reverse phase chromatography on a Vydac C4 column 4.6 x 150 mm (Western Analytical, Temecula, CA) with a linear gradient of 0-100% acetonitrile (0.1% TFA) at a flow rate of 2 ml/min. Eluent was monitored at 214 nm

The fraction with biological activity was separated into one major peak plus other smaller components. Western blot of this peak onto PVDF membrane revealed three major bands of molecular weights 14 Kd. 20 Kd and 26 Kd. These polypeptides were determined to have the following N-terminal sequences, respectively:

(j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 134)

- (k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID No. 135) and
- (l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136), wherein Xaa may be any amino acid.
- Using the assays described above, these polypeptides were shown to induce proliferation and IFN-γ production in PBMC preparations. Figs. 1A and B show the results of such assays using PBMC preparations from a first and a second donor, respectively.

DNA sequences that encode the antigens designated as (a), (c), (d) and (g) above were obtained by screening a genomic *M. tuberculosis* library using ³²P end labeled degenerate oligonucleotides corresponding to the N-terminal sequence and containing *M. tuberculosis* codon bias. The screen performed using a probe corresponding to antigen (a) above identified a clone having the sequence provided in SEQ ID No. 101. The polypeptide encoded by SEQ ID No. 101 is provided in SEQ ID No. 102. The screen performed using a probe corresponding to antigen (g) above identified a clone having the sequence provided in SEQ ID No. 52. The polypeptide encoded by SEQ ID No. 52 is provided in SEQ ID No. 53. The screen performed using a probe corresponding to antigen (d) above identified a clone having the sequence provided in SEQ ID No. 24. and the screen performed with a probe corresponding to antigen (c) identified a clone having the sequence provided in SEQ ID No: 25.

The above amino acid sequences were compared to known amino acid sequences in the gene bank using the DNA STAR system. The database searched contains some 173,000 proteins and is a combination of the Swiss, PIR databases along with translated protein sequences (Version 87). No significant homologies to the amino acid sequences for antigens (a)-(h) and (l) were detected.

The amino acid sequence for antigen (i) was found to be homologous to a sequence from *M. leprae*. The full length *M. leprae* sequence was amplified from genomic DNA using the sequence obtained from GENBANK. This sequence was then used to screen the *M. tuberculosis* library described below in Example 2 and a full length copy of the *M. tuberculosis* homologue was obtained (SEQ ID No. 99).

The amino acid sequence for antigen (j) was found to be homologous to a known *M. tuberculosis* protein translated from a DNA sequence. To the best of the inventors' knowledge, this protein has not been previously shown to possess T-cell stimulatory activity. The amino acid sequence for antigen (k) was found to be related to a sequence from *M. leprae*.

In the proliferation and IFN- γ assays described above, using three PPD positive donors, the results for representative antigens provided above are presented in Table 1:

TABLE 1

RESULTS OF PBMC PROLIFERATION AND IFN-y ASSAYS

Sequence	Proliferation	IFN-γ
(a)	+	
(c)	+++	
(d)	++	+++
(g)	+++	**
(h)	+++	+++

In Table 1, responses that gave a stimulation index (SI) of between 2 and 4 (compared to cells cultured in medium alone) were scored as +, an SI of 4-8 or 2-4 at a concentration of 1 µg or less was scored as ++ and an SI of greater than 8 was scored as +++. The antigen of sequence (i) was found to have a high SI (++++) for one donor and lower SI (+++ and +) for the two other donors in both proliferation and IFN-y assays. These results indicate that these antigens are capable of inducing proliferation and/or interferon-y production.

30

EXAMPLE 2

USE OF PATIENT SERA TO ISOLATE M. TUBERCULOSIS ANTIGENS

This example illustrates the isolation of antigens from *M. tuberculosis*lysate by screening with serum from *M. tuberculosis*-infected individuals.

Dessicated *M. tuberculosis* H37Ra (Difco Laboratories) was added to a 2% NP40 solution, and alternately homogenized and sonicated three times. The resulting suspension was centrifuged at 13,000 rpm in microfuge tubes and the supernatant put through a 0.2 micron syringe filter. The filtrate was bound to Macro Prep DEAE beads (BioRad, Hercules, CA). The beads were extensively washed with 20 mM Tris pH 7.5 and bound proteins eluted with 1M NaCl. The 1M NaCl elute was dialyzed overnight against 10 mM Tris, pH 7.5. Dialyzed solution was treated with DNase and RNase at 0.05 mg/ml for 30 min. at room temperature and then with α-D-mannosidase, 0.5 U/mg at pH 4.5 for 3-4 hours at room temperature. After returning to pH 7.5, the material was fractionated via FPLC over a Bio Scale-Q-20 column (BioRad). Fractions were combined into nine pools, concentrated in a Centriprep 10 (Amicon, Beverley, MA) and then screened by Western blot for serological activity using a serum pool from *M. tuberculosis*-infected patients which was not immunoreactive with other antigens of the present invention.

- The most reactive fraction was run in SDS-PAGE and transferred to PVDF. A band at approximately 85 Kd was cut out yielding the sequence:
 - (m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No. 137), wherein Xaa may be any amino acid.
- Comparison of this sequence with those in the gene bank as described above, revealed no significant homologies to known sequences.

A DNA sequence that encodes the antigen designated as (m) above was obtained by screening a genomic M. tuberculosis Erdman strain library using labeled degenerate oligonucleotides corresponding to the N-terminal sequence of SEQ ID NO:137. A clone was identified having the DNA sequence provided in SEQ ID NO:

203. This sequence was found to encode the amino acid sequence provided in SEQ ID NO: 204. Comparison of these sequences with those in the genebank revealed some similarity to sequences previously identified in *M. tuberculosis* and *M. bovis*.

5

10

15

30

EXAMPLE 3

PREPARATION OF DNA SEQUENCES ENCODING M. TUBERCULOSIS ANTIGENS

This example illustrates the preparation of DNA sequences encoding *M. tuberculosis* antigens by screening a *M. tuberculosis* expression library with sera obtained from patients infected with *M. tuberculosis*, or with anti-sera raised against soluble *M. tuberculosis* antigens.

A. PREPARATION OF M. TUBERCULOSIS SOLUBLE ANTIGENS USING RABBIT ANTISERA RAISED AGAINST M. TUBERCULOSIS SUPERNATANT

Genomic DNA was isolated from the *M. tuberculosis* strain H37Ra. The DNA was randomly sheared and used to construct an expression library using the Lambda ZAP expression system (Stratagene, La Jolla, CA). Rabbit anti-sera was generated against secretory proteins of the *M. tuberculosis* strains H37Ra, H37Rv and Erdman by immunizing a rabbit with concentrated supernatant of the *M. tuberculosis* cultures. Specifically, the rabbit was first immunized subcutaneously with 200 µg of protein antigen in a total volume of 2 ml containing 10 µg muramyl dipeptide (Calbiochem, La Jolla, CA) and 1 ml of incomplete Freund's adjuvant. Four weeks later the rabbit was boosted subcutaneously with 100 µg antigen in incomplete Freund's adjuvant. Finally, the rabbit was immunized intravenously four weeks later with 50 µg protein antigen. The anti-sera were used to screen the expression library as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratories. Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones deduced.

Thirty two clones were purified. Of these, 25 represent sequences that have not been previously identified in human *M. tuberculosis*. Recombinant antigens

15

were expressed and purified antigens used in the immunological analysis described in Example 1. Proteins were induced by IPTG and purified by gel elution, as described in Skeiky et al., *J. Exp. Med.* 181:1527-1537, 1995. Representative sequences of DNA molecules identified in this screen are provided in SEQ ID Nos.: 1-25. The corresponding predicted amino acid sequences are shown in SEQ ID Nos. 63-87.

On comparison of these sequences with known sequences in the gene bank using the databases described above, it was found that the clones referred to hereinafter as TbRA2A, TbRA16, TbRA18, and TbRA29 (SEQ ID Nos. 76, 68, 70, 75) show some homology to sequences previously identified in *Mycobacterium leprae* but not in *M. tuberculosis*. TbRA2A was found to be a lipoprotein, with a six residue lipidation sequence being located adjacent to a hydrophobic secretory sequence. TbRA11, TbRA26, TbRA28 and TbDPEP (SEQ ID Nos.: 65, 73, 74, 53) have been previously identified in *M. tuberculosis*. No significant homologies were found to TbRA1, TbRA3. TbRA4, TbRA9, TbRA10, TbRA13, TbRA17, TbRa19, TbRA29, TbRA32, TbRA36 and the overlapping clones TbRA35 and TbRA12 (SEQ ID Nos. 63, 77, 81, 82, 64, 67, 69, 71, 75, 78, 80, 79, 66). The clone TbRa24 is overlapping with clone TbRa29.

The results of PBMC proliferation and interferon- γ assays performed on representative recombinant antigens, and using T-cell preparations from several different *M. tuberculosis*-immune patients, are presented in Tables 2 and 3, respectively.

<u>TABLE 2</u> Results of PBMC Proliferation To Representative Soluble Antigens

					Τ	T	T	T	T	Т	_			Т	\neg	-		_		T -				_
			13	•	'						E	=	1		ᄅ	Ξ	ä	=	ŭ	=	:	≡	=	
			2	++	<u> </u>	i	+	1	+	•	Ē	ī	T		=	‡	E	II.	ĭ	+	1	- -	H	,
			=	+		Ē	++	‡	:	+	ē	ם ב	į	1	E	‡	Ħ	Ξ	nt	111	1	1	+	
			2	ı	,	ä	+	‡		•	Ē	Ħ	=	:	=	‡	=	E	Ξ	,		+	-	
		ĺ	7		#	nt	'	±	+	4 3	=	=	Ξ	1	=	<u>+</u>	=	=	E	,	,	+	,	
		·	•	+1	++	ju	#		+	1 3	I	=	Ħ	Ξ			ĕ	ĕ	Ξ	•	•	‡		•
	Patient	7		+1	•	nt	nt	Ĕ	+	=		=	Ħ	E		= ;	= ;	=	E	E	=	=		,
		٧	,		,	++	+	+	+	+			•	•	‡		<u> </u>			-	,			
		5		•	#	+	#	‡	+					•	<u>+</u>					•	,			
		4	-			Ħ	#	-	÷	Ħ	=		=	Ξ	÷	=	1	1	=		•	+		
		3	+	1	+	II	++	-	+	H	=		=	nt	Ī	=	1	1		#	-	,		1
		2			н	'	,	+1	•	Ξ	Ħ	-	-	Ξ	Ξ	E	Ξ	E		'	'	+	,	
		-	,		•	•	•	+1		nt	111			Ē	=	Ħ	E	=		•	•	•		
Antigen			TbRal	ThRai	G	1 bKa9	IbRa10	Ibkall	TbRa12	TbRa16	lbRa24	ThRack	2	1bRa29	TbRa35	TbRaB	TbRaC	TbRaD	AAAAV	4	٨٨	DPEP	Control	
				-								<u> </u>			لت					٠ .		<u> </u>)	-

nt = not tested

	1	ı		- 1	-	- 1					Τ	\Box				Г	Т	T	1	Т	7	_	_
			=	_'		Ē	_	ı		Ē		Ħ	Ξ	ē	ī	=	=	: :	=	=	Ξ	Ξ	•
IGENS			12	++		ĕ	+	+1		10		E	10	nt	‡	ĭ	1	1	=	#	+	+	•
ILE ANT			=	+	•	n	+1	‡	+	nt	1	≅	Ξ	Ξ	‡ ‡ ‡	nt	10	1	=	E	E	#1	,
E SOLUB			9		•	Ξ	+	++	,	'n	1	=	ē	Ξ	‡	Ξ	Ξ	1		,	+	÷	
NTATIV		,	6		++	Ξ	•	+	++	=	=		=	E	÷ +	Ħ	ĭ	=				41	
REPRESE				#1	‡	=	#		#	Ħ	=	1		=	‡	Ξ	Ξ	ī			•	+	•
RESULTS OF PRIMC INTERFERON-Y PRODUCTION TO REPRESENTATIVE SOLUBLE ANTIGENS	Patient	-				=	ij	ıı	+	ıı	u	à		=	E	=	=	Ξ	=	1	= -	Ĕ	
TAB		_				, .	+	+	++++	+		+			+	+	+	+			+		•
KON-Y P		,	, -	- +	1 ‡	: -	н :	÷]	+1	+	+	+	-			‡	+	÷			-	-	•
NTERFE		7	=		1	+	4		+	Ξ	Ħ	=	1		-	=	Ξ	nt			1 1 1	+	•
PBMC		_		‡	1	+	4	-	-	E	Ξ	7	=	1	: ;	=	E	=	#		+		•
JUS OF		2	Ξ	++	+	-	+	+		=	Ξ	+	Ξ	=		=	=	Ξ	,		-	-	-
REST		_	+		+ +	+				=	Ξ	+	Ξ	7	=		=	Ξ	,	,	+		
į	Antigen		TbRai	TbRa3	TbRa9	TbRaio	FbRall	ThRail	7.40.17	OKATO	TbRa24	ľbRa26	TbRa29	TbRa35	ThRail	T. P.	UNAC	TbRaD	AAMK	٠. ۲.	DPEP	Control	20111102
Į	∢		1.b	ء: _	۾ ا	1.b	2	[1	0		rb	TrbR	TbR	T.P.R	1.1	-	Tok	AA	λλ	DPE	0	

15

20

25

In Tables 2 and 3, responses that gave a stimulation index (SI) of between 1.2 and 2 (compared to cells cultured in medium alone) were scored as ±, a SI of 2-4 was scored as +, as SI of 4-8 or 2-4 at a concentration of 1 µg or less was scored as ++ and an SI of greater than 8 was scored as +++. In addition, the effect of concentration on proliferation and interferon-y production is shown for two of the above antigens in the attached Figure. For both proliferation and interferon-y production. TbRa3 was scored as ++ and TbRa9 as +.

These results indicate that these soluble antigens can induce proliferation and/or interferon-y production in T-cells derived from an *M. tuberculosis*-immune individual.

B. USE OF SERA FROM PATIENTS HAVING PULMONARY OR PLEURAL TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING M. TUBERCULOSIS ANTIGENS

The genomic DNA library described above, and an additional H37Rv library, were screened using pools of sera obtained from patients with active tuberculosis. To prepare the H37Rv library, *M. tuberculosis* strain H37Rv genomic DNA was isolated, subjected to partial *Sau*3A digestion and used to construct an expression library using the Lambda Zap expression system (Stratagene, La Jolla, Ca). Three different pools of sera, each containing sera obtained from three individuals with active pulmonary or pleural disease, were used in the expression screening. The pools were designated TbL, TbM and TbH, referring to relative reactivity with H37Ra lysate (*i.e.*, TbL = low reactivity. TbM = medium reactivity and TbH = high reactivity) in both ELISA and immunoblot format. A fourth pool of sera from seven patients with active pulmonary tuberculosis was also employed. All of the sera lacked increased reactivity with the recombinant 38 kD *M. tuberculosis* H37Ra phosphate-binding protein.

All pools were pre-adsorbed with *E. coli* lysate and used to screen the H37Ra and H37Rv expression libraries, as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones deduced.

WO 99/42076 PCT/US99/03268

Thirty two clones were purified. Of these, 31 represented sequences that had not been previously identified in human *M. tuberculosis*. Representative sequences of the DNA molecules identified are provided in SEQ ID Nos.: 26-51 and 105. Of these, TbH-8-2 (SEQ. ID NO. 105) is a partial clone of TbH-8, and TbH-4 (SEQ. ID NO. 43) and TbH-4-FWD (SEQ. ID NO. 44) are non-contiguous sequences from the same clone. Amino acid sequences for the antigens hereinafter identified as Tb38-1. TbH-4, TbH-8, TbH-9, and TbH-12 are shown in SEQ ID Nos.: 88-92. Comparison of these sequences with known sequences in the gene bank using the databases identified above revealed no significant homologies to TbH-4, TbH-8, TbH-9 and TbM-3, although weak homologies were found to TbH-9. TbH-12 was found to be homologous to a 34 kD antigenic protein previously identified in *M. paratuberculosis* (Acc. No. S28515). Tb38-1 was found to be located 34 base pairs upstream of the open reading frame for the antigen ESAT-6 previously identified in *M. bovis* (Acc. No. U34848) and in *M. tuberculosis* (Sorensen et al., *Infec. Immun. 63*:1710-1717, 1995).

Probes derived from Tb38-1 and TbH-9, both isolated from an H37Ra library, were used to identify clones in an H37Rv library. Tb38-1 hybridized to Tb38-1F2. Tb38-1F3. Tb38-1F5 and Tb38-1F6 (SEQ. ID NOS. 112, 113, 116, 118, and 119). (SEQ ID NOS. 112 and 113 are non-contiguous sequences from clone Tb38-1F2.) Two open reading frames were deduced in Tb38-IF2; one corresponds to Tb37FL (SEQ. ID. NO. 114), the second, a partial sequence, may be the homologue of Tb38-1 and is called Tb38-IN (SEQ. ID NO. 115). The deduced amino acid sequence of Tb38-1F3 is presented in SEQ. ID. NO. 117. A TbH-9 probe identified three clones in the H37Rv library: TbH-9-FL (SEQ. ID NO. 106), which may be the homologue of TbH-9 (R37Ra), TbH-9-1 (SEQ. ID NO. 108), and TbH-9-4 (SEQ. ID NO. 110), all of which are highly related sequences to TbH-9. The deduced amino acid sequences for these three clones are presented in SEQ ID NOS. 107, 109 and 111.

15

30

Further screening of the *M. tuberculosis* genomic DNA library. as described above, resulted in the recovery of ten additional reactive clones, representing seven different genes. One of these genes was identified as the 38 Kd antigen discussed

WO 99/42076 PCT/US99/03268

49

above, one was determined to be identical to the 14Kd alpha crystallin heat shock protein previously shown to be present in *M. tuberculosis*, and a third was determined to be identical to the antigen TbH-8 described above. The determined DNA sequences for the remaining five clones (hereinafter referred to as TbH-29, TbH-30, TbH-32 and TbH-33) are provided in SEQ ID NO: 138-141, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 142-145, respectively. The DNA and amino acid sequences for these antigens were compared with those in the gene bank as described above. No homologies were found to the 5' end of TbH-29 (which contains the reactive open reading frame), although the 3' end of TbH-29 was found to be identical to the *M. tuberculosis* cosmid Y227. TbH-32 and TbH-33 were found to be identical to the previously identified *M. tuberculosis* insertion element IS6110 and to the *M. tuberculosis* cosmid Y50, respectively. No significant homologies to TbH-30 were found.

10

Positive phagemid from this additional screening were used to infect E. coli XL-1 Blue MRF', as described in Sambrook et al., supra. Induction of recombinant protein was accomplished by the addition of IPTG. Induced and uninduced lysates were run in duplicate on SDS-PAGE and transferred to nitrocellulose filters. Filters were reacted with human M. tuberculosis sera (1:200 dilution) reactive with TbH and a rabbit sera (1:200 or 1:250 dilution) reactive with the N-terminal 4 Kd portion of lacZ. Sera incubations were performed for 2 hours at room temperature. Bound antibody was detected by addition of ¹²⁵I-labeled Protein A and subsequent exposure to film for variable times ranging from 16 hours to 11 days. The results of the immunoblots are summarized in Table 4.

TABLE 4

5	Antigen	Human M. tb <u>Sera</u>	Anti-lacZ <u>Sera</u>
	Тън-29	45 Kd	45 Kd
	Тън-30	No reactivity	29 Kd
	Тьн-32	12 Kd	12 Kd
	Тьн-33	16 Kd	16 Kd

25

Positive reaction of the recombinant human *M. tuberculosis* antigens with both the human *M. tuberculosis* sera and anti-lacZ sera indicate that reactivity of the human *M. tuberculosis* sera is directed towards the fusion protein. Antigens reactive with the anti-lacZ sera but not with the human *M. tuberculosis* sera may be the result of the human *M. tuberculosis* sera recognizing conformational epitopes, or the antigen-antibody binding kinetics may be such that the 2 hour sera exposure in the immunoblot is not sufficient.

The results of T-cell assays performed on Tb38-1, ESAT-6 and other representative recombinant antigens are presented in Tables 5A, B and 6, respectively, below:

TABLE 5A

RESULTS OF PBMC PROLIFERATION TO REPRESENTATIVE ANTIGENS

Antigen						Donor					
	ı	2	3	4	5	6	7	8	0	10	
Tb38.1		-	-		-	+		+		10	11
ESAT-6		+	-	+_	_	+		<u> </u>	-		
Тън-9	- 1	++		-	<u>+</u>	+	++	 	+	++	

10

15

20

TABLE 5B

RESULTS OF PBMC INTERFERON-Y PRODUCTION TO REPRESENTATIVE ANTIGENS

Antigen	-					Donor					
	1	2	3	4	5	6	7	8	9	10	11
Tb38.1	+++	+	-	+	+	+++	-	++	-	+++	+++
ESAT-6	+	+	+	+	+-	+	-	+	+	+	
Тън-9	++	++	-	+++	±	±	+++	+++	++		+-

TABLE 6
SUMMARY OF T-CELL RESPONSES TO REPRESENTATIVE ANTIGENS

		roliferation	n.		Interferon-	Y	
Antigen	patient 4	patient 5	patient 6	patient 4	patient 5	patient 6	total
Тън9	++	++	++	+++	++	++	13
Тьм7	-	÷	-	+-	+	-	4
Тън5	-	+	+	++	++	++	8
TbL23	-	+	±	++	++	+	7.5
Тън4	-	++	±	++	++	±	7
- control	-	-	-	-	-	-	0

These results indicate that both the inventive *M. tuberculosis* antigens and ESAT-6 can induce proliferation and/or interferon-y production in T-cells derived from an *M. tuberculosis*-immune individual. To the best of the inventors' knowledge. ESAT-6 has not been previously shown to stimulate human immune responses

A set of six overlapping peptides covering the amino acid sequence of the antigen Tb38-1 was constructed using the method described in Example 6. The sequences of these peptides, hereinafter referred to as pep1-6, are provided in SEQ ID Nos. 93-98, respectively. The results of T-cell assays using these peptides are shown in Tables 7 and 8. These results confirm the existence, and help to localize T-cell epitopes within Tb38-1 capable of inducing proliferation and interferon-γ production in T-cells derived from an *M. tuberculosis* immune individual.

TABLE 7 RESULTS OF PBMC PROLIFERATION TO TB38-1 PEPTIDES

Peplide							Patient						
	_	2	3	4	2	9	7	∞	6	01	=	12	13
pep1			,		+1		٠	•	,	H	•	•	+
pep2	+1	,		•	+1	,	•	•	+1	++	•	•	+
pep3		,				,	•	,	+4	•	•		#
pep4	Ξ	<u> </u>					+		#1	##	•		+
pep5	+	+1	,			,	+	,	#1	•	-		+
9dad		‡	•	,		,	#	•	+1	+	•		+
Control	,		,	,	,	,	•	1	1	,	•		'

TABLE 8 RESULTS OF PBMC INTERFERON-Y PRODUCTION TO TB38-1 PEPTIDES

		=	2 +	. +	- .	#	+	+	+	
		12					,	,		
31		=		1.		'	•			
		01	++	+		•	#	•	+	
		6		#1	+	4	#	#1	÷i	
		∞						,		
	Patient	7			,		÷	+	+1	
		9						,	,	
		5	#	++				•	•	•
		7			,			,	•	
		3			,					
		2		'				+1	-	,
		-	-		,	:		-	<u>-</u>	
	Peptide		pepi	pep2	pep3	Pen-4		pep5	9dad	Control
				_						

Studies were undertaken to determine whether the antigens TbH-9 and Tb38-1 represent cellular proteins or are secreted into *M. tuberculosis* culture media. In the first study, rabbit sera were raised against A) secretory proteins of *M. tuberculosis*, B) the known secretory recombinant *M. tuberculosis* antigen 85b, C) recombinant Tb38-1 and D) recombinant TbH-9, using protocols substantially the same as that as described in Example 3A. Total *M. tuberculosis* lysate, concentrated supernatant of *M. tuberculosis* cultures and the recombinant antigens 85b, TbH-9 and Tb38-1 were resolved on denaturing gels, immobilized on nitrocellulose membranes and duplicate blots were probed using the rabbit sera described above.

The results of this analysis using control sera (panel I) and antisera (panel II) against secretory proteins, recombinant 85b, recombinant Tb38-1 and recombinant TbH-9 are shown in Figures 3A-D, respectively, wherein the lane designations are as follows: 1) molecular weight protein standards; 2) 5 µg of *M. tuberculosis* lysate; 3) 5 µg secretory proteins: 4) 50 ng recombinant Tb38-1; 5) 50 ng recombinant TbH-9; and 6) 50 ng recombinant 85b. The recombinant antigens were engineered with six terminal histidine residues and would therefore be expected to migrate with a mobility approximately 1 kD larger that the native protein. In Figure 3D, recombinant TbH-9 is lacking approximately 10 kD of the full-length 42 kD antigen, hence the significant difference in the size of the immunoreactive native TbH-9 antigen in the lysate lane (indicated by an arrow). These results demonstrate that Tb38-1 and TbH-9 are intracellular antigens and are not actively secreted by *M. tuberculosis*.

The finding that TbH-9 is an intracellular antigen was confirmed by determining the reactivity of TbH-9-specific human T cell clones to recombinant TbH-9. secretory *M. tuberculosis* proteins and PPD. A TbH-9-specific T cell clone (designated 131TbH-9) was generated from PBMC of a healthy PPD-positive donor. The proliferative response of 131TbH-9 to secretory proteins, recombinant TbH-9 and a control *M. tuberculosis* antigen. TbRall, was determined by measuring uptake of tritiated thymidine, as described in Example 1. As shown in Figure 4A, the clone 131TbH-9 responds specifically to TbH-9, showing that TbH-9 is not a significant component of *M. tuberculosis* secretory proteins. Figure 4B shows the production of IFN-y by a second TbH-9-specific T cell clone

(designated PPD 800-10) prepared from PBMC from a healthy PPD-positive donor, following stimulation of the T cell clone with secretory proteins, PPD or recombinant TbH-9. These results further confirm that TbH-9 is not secreted by M. tuberculosis.

C. USE OF SERA FROM PATIENTS HAVING EXTRAPULMONARY TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING M. TUBERCULOSIS ANTIGENS

Genomic DNA was isolated from M. tuberculosis Erdman strain, randomly sheared and used to construct an expression library employing the Lambda ZAP expression system (Stratagene. La Jolla, CA). The resulting library was screened using pools of sera obtained from individuals with extrapulmonary tuberculosis, as described above in Example 3B, with the secondary antibody being goat anti-human IgG + A + M (H+L) conjugated with alkaline phosphatase.

Eighteen clones were purified. Of these, 4 clones (hereinafter referred to as XP14, XP24, XP31 and XP32) were found to bear some similarity to known sequences. The determined DNA sequences for XP14, XP24 and XP31 are provided in SEQ ID Nos.: 156-158, respectively, with the 5' and 3' DNA sequences for XP32 being provided in SEQ ID Nos.: 159 and 160, respectively. The predicted amino acid sequence for XP14 is provided in SEQ ID No: 161. The reverse complement of XP14 was found to encode the amino acid sequence provided in SEQ ID No.: 162.

Comparison of the sequences for the remaining 14 clones (hereinafter referred to as XP1-XP6. XP17-XP19. XP22. XP25. XP27. XP30 and XP36) with those in the genebank as described above, revealed no homologies with the exception of the 3' ends of XP2 and XP6 which were found to bear some homology to known *M. tuberculosis* cosmids. The DNA sequences for XP27 and XP36 are shown in SEQ ID Nos.: 163 and 164, respectively, with the 5' sequences for XP4, XP5, XP17 and XP30 being shown in SEQ ID Nos: 165-168, respectively, and the 5' and 3' sequences for XP2. XP3, XP6, XP18, XP19, XP22 and XP25 being shown in SEQ ID Nos: 169 and 170; 171 and 172; 173 and 174: 175 and 176; 177 and 178; 179 and 180; and 181 and 182, respectively. XP1 was found to overlap with the DNA sequences for TbH4, disclosed above. The full-length DNA sequence for TbH4-XP1 is provided in SEQ ID No.: 183. This DNA sequence was found to contain an

open reading frame encoding the amino acid sequence shown in SEQ ID No: 184. The reverse complement of TbH4-XP1 was found to contain an open reading frame encoding the amino acid sequence shown in SEQ ID No.: 185. The DNA sequence for XP36 was found to contain two open reading frames encoding the amino acid sequence shown in SEQ ID Nos.: 186 and 187, with the reverse complement containing an open reading frame encoding the amino acid sequence shown in SEQ ID No.: 188.

Recombinant XP1 protein was prepared as described above in Example 3B, with a metal ion affinity chromatography column being employed for purification. As illustrated in Figures 8A-B and 9A-B, using the assays described herein, recombinant XP1 was found to stimulate cell proliferation and IFN- γ production in T cells isolated from an M tuberculosis-immune donors.

D. USE OF A LYSATE POSITIVE SERUM POOL FROM PATIENTS HAVING TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING M. TUBERCULOSIS ANTIGENS

Genomic DNA was isolated from M. tuberculosis Erdman strain, randomly sheared and used to construct an expression library employing the Lambda Screen expression system (Novagen, Madison, WI), as described below in Example 6. Pooled serum obtained from M. tuberculosis-infected patients and that was shown to react with M. tuberculosis lysate but not with the previously expressed proteins 38kD, Tb38-1, TbRa3, TbH4, DPEP and TbRa11, was used to screen the expression library as described above in Example 3B, with the secondary antibody being goat anti-human IgG + A + M (H+L) conjugated with alkaline phosphatase.

Twenty-seven clones were purified. Comparison of the determined cDNA sequences for these clones revealed no significant homologies to 10 of the clones (hereinafter referred to as LSER-10, LSER-11, LSER-12, LSER-13, LSER-16, LSER-18, LSER-23, LSER-24, LSER-25 and LSER-27). The determined 5' cDNA sequences for LSER-10, LSER-11, LSER-12, LSER-13, LSER-16 and LSER-25 are provided in SEQ ID NO: 242-247, respectively, with the corresponding predicted amino acid sequences for LSER-10, LSER-12, LSER-13, LSER-16 and LSER-25 being provided in SEQ ID NO: 248-252, respectively. The determined full-length cDNA sequences for LSER-18, LSER-23, LSER-24

and LSER-27 are shown in SEQ ID NO: 253-256, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 257-260. The remaining seventeen clones were found to show similarities to unknown sequences previously identified in *M. tuberculosis*. The determined 5' cDNA sequences for sixteen of these clones (hereinafter referred to as LSER-1, LSER-3, LSER-4, LSER-5, LSER-6, LSER-8, LSER-14, LSER-15, LSER-17, LSER-19, LSER-20, LSER-22, LSER-26, LSER-28, LSER-29 and LSER-30) are provided in SEQ ID NO: 261-276, respectively, with the corresponding predicted amino acid sequences for LSER-1, LSER-3, LSER-5, LSER-6, LSER-8, LSER-14, LSER-15, LSER-17, LSER-19, LSER-20, LSER-22, LSER-26, LSER-28, LSER-29 and LSER-30 being provided in SEQ ID NO: 277-291, respectively. The determined full-length cDNA sequence for the clone LSER-9 is provided in SEQ ID NO: 292. The reverse complement of LSER-6 (SEQ ID NO: 293) was found to encode the predicted amino acid sequence of SEQ ID NO: 294.

E. PREPARATION OF M. TUBERCULOSIS SOLUBLE ANTIGENS USING RABBIT ANTI-SERA RAISED AGAINST M. TUBERCULOSIS FRACTIONATED PROTEINS

M. tuberculosis lysate was prepared as described above in Example 2. The resulting material was fractionated by HPLC and the fractions screened by Western blot for serological activity with a serum pool from M. tuberculosis-infected patients which showed little or no immunoreactivity with other antigens of the present invention. Rabbit anti-sera was generated against the most reactive fraction using the method described in Example 3A. The anti-sera was used to screen an M. tuberculosis Erdman strain genomic DNA expression library prepared as described above. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the M. tuberculosis clones determined.

Ten different clones were purified. Of these, one was found to be TbRa35, described above, and one was found to be the previously identified *M. tuberculosis* antigen. HSP60. Of the remaining eight clones, seven (hereinafter referred to as RDIF2, RDIF5, RDIF8, RDIF10, RDIF11 and RDIF 12) were found to bear some similarity to previously identified *M. tuberculosis* sequences. The determined DNA sequences for RDIF2, RDIF5, RDIF8, RDIF10 and RDIF11 are provided in SEQ ID Nos.: 189-193, respectively, with the

corresponding predicted amino acid sequences being provided in SEQ ID Nos: 194-198, respectively. The 5' and 3' DNA sequences for RDIF12 are provided in SEQ ID Nos.: 199 and 200, respectively. No significant homologies were found to the antigen RDIF-7. The determined DNA and predicted amino acid sequences for RDIF7 are provided in SEQ ID Nos.: 201 and 202, respectively. One additional clone, referred to as RDIF6 was isolated, however, this was found to be identical to RDIF5.

Recombinant RDIF6. RDIF8, RDIF10 and RDIF11 were prepared as described above. As shown in Figures 8A-B and 9A-B, these antigens were found to stimulate cell proliferation and IFN-y production in T cells isolated from M. tuberculosis-immune donors.

EXAMPLE 4

PURIFICATION AND CHARACTERIZATION OF A POLYPEPTIDE FROM TUBERCULIN PURIFIED PROTEIN DERIVATIVE

An M. tuberculosis polypeptide was isolated from tuberculin purified protein derivative (PPD) as follows.

PPD was prepared as published with some modification (Seibert, F. et al., Tuberculin purified protein derivative. Preparation and analyses of a large quantity for standard. The American Review of Tuberculosis 44:9-25, 1941).

M. tuberculosis Rv strain was grown for 6 weeks in synthetic medium in roller bottles at 37°C. Bottles containing the bacterial growth were then heated to 100° C in water vapor for 3 hours. Cultures were sterile filtered using a 0.22 μ filter and the liquid phase was concentrated 20 times using a 3 kD cut-off membrane. Proteins were precipitated once with 50% ammonium sulfate solution and eight times with 25% ammonium sulfate solution. The resulting proteins (PPD) were fractionated by reverse phase liquid chromatography (RP-HPLC) using a C18 column (7.8 x 300 mM: Waters. Milford. MA) in a Biocad HPLC system (Perseptive Biosystems. Framingham, MA). Fractions were eluted from the column with a linear gradient from 0-100% buffer (0.1% TFA in acetonitrile). The flow rate was 10 ml/minute and eluent was monitored at 214 nm and 280 nm.

Six fractions were collected, dried, suspended in PBS and tested individually in *M. tuberculosis*-infected guinea pigs for induction of delayed type hypersensitivity (DTH) reaction. One fraction was found to induce a strong DTH reaction and was subsequently fractionated further by RP-HPLC on a microbore Vydac C18 column (Cat. No. 218TP5115) in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted with a linear gradient from 5-100% buffer (0.05% TFA in acetonitrile) with a flow rate of 80 µl/minute. Eluent was monitored at 215 nm. Eight fractions were collected and tested for induction of DTH in *M. tuberculosis*-infected guinea pigs. One fraction was found to induce strong DTH of about 16 mm induration. The other fractions did not induce detectable DTH. The positive fraction was submitted to SDS-PAGE gel electrophoresis and found to contain a single protein band of approximately 12 kD molecular weight.

This polypeptide, herein after referred to as DPPD, was sequenced from the amino terminal using a Perkin Elmer/Applied Biosystems Division Procise 492 protein sequencer as described above and found to have the N-terminal sequence shown in SEQ ID No.: 129. Comparison of this sequence with known sequences in the gene bank as described above revealed no known homologies. Four cyanogen bromide fragments of DPPD were isolated and found to have the sequences shown in SEQ ID Nos.: 130-133. A subsequent search of the *M. tuberculosis* genome database released by the Institute for Genomic Research revealed a match of the DPPD partial amino acid sequence with a sequence present within the *M. tuberculosis* cosmid MTY21C12. An open reading frame of 336 bp was identified. The full-length DNA sequence for DPPD is provided in SEQ ID NO: 240. with the corresponding full-length amino acid sequence being provided in SEQ ID NO: 241.

The ability of the antigen DPPD to stimulate human PBMC to proliferate and to produce IFN-γ was assayed as described in Example 1. As shown in Table 9, DPPD was found to stimulate proliferation and elicit production of large quantities of IFN-γ; more than that elicited by commercial PPD.

TABLE 9

RESULTS OF PROLIFERATION AND INTERFERON-y ASSAYS TO DPPD

PBMC Donor	Stimulator	Proliferation (CPM)	IFN-γ (OD ₁₅₀)
<u>A</u>	Medium	1,089	0.17
	PPD (commercial)	8,394	1.29
	DPPD	13,451	2.21
<u> </u>			
В	Medium	450	0.09
	PPD (commercial)	3,929	1.26
	DPPD	6,184	1.49
C	N (- 1'		
	Medium	541	0.11
	PPD (commercial)	8,907	0.76
	DPPD	23,024	>2.70

EXAMPLE 5 Use of Sera From Tuberculosis-Infected Monkeys to Identify DNA Sequences Encoding M. Tuberculosis Antigens

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda ZAP expression system (Stratagene, La Jolla, CA). Serum samples were obtained from a cynomolgous monkey 18, 33, 51 and 56 days following infection with *M. tuberculosis* Erdman strain. These samples were pooled and used to screen the *M. tuberculosis* genomic DNA expression library using the procedure described above in Example 3C.

Twenty clones were purified. The determined 5' DNA sequences for the clones referred to as MO-1, MO-2, MO-4, MO-8, MO-9, MO-26, MO-28, MO-29, MO-30, MO-34 and MO-35 are provided in SEQ ID NO: 215-225, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 226-236. The full-length DNA sequence of the clone MO-10 is provided in SEQ ID NO: 237, with the corresponding

predicted amino acid sequence being provided in SEQ ID NO: 238. The 3' DNA sequence for the clone MO-27 is provided in SEQ ID NO: 239.

Clones MO-1. MO-30 and MO-35 were found to show a high degree of relatedness and showed some homology to a previously identified unknown *M. tuberculosis* sequence and to cosmid MTCI237. MO-2 was found to show some homology to aspartokinase from *M. tuberculosis*. Clones MO-3, MO-7 and MO-27 were found to be identical and to show a high degree of relatedness to MO-5. All four of these clones showed some homology to *M. tuberculosis* heat shock protein 70. MO-27 was found to show some homology to *M. tuberculosis* cosmid MTCY339. MO-4 and MO-34 were found to show some homology to cosmid SCY21B4 and *M. smegmatis* integration host factor, and were both found to show some homology to a previously identified, unknown *M. tuberculosis* sequence. MO-6 was found to show some homology to *M. tuberculosis* heat shock protein 65. MO-8, MO-9, MO-10, MO-26 and MO-29 were found to be highly related to each other and to show some homology to *M. tuberculosis* dihydrolipamide succinyltransferase. MO-28, MO-31 and MO-32 were found to be identical and to show some homology to a previously identified *M. tuberculosis* protein. MO-33 was found to show some homology to a previously identified *M. tuberculosis* protein. MO-33 was found to show some homology to a previously identified *M. tuberculosis* protein. MO-33 was found to show some homology to a

Further studies using the above protocol resulted in the isolation of an additional four clones, hereinarter referred to as MO-12, MO-13, MO-19 and MO-39. The determined 5° cDNA sequences for these clones are provided in SEQ ID NO: 295-298, respectively, with the corresponding predicted protein sequences being provided in SEQ ID NO: 299-302, respectively. Comparison of these sequences with those in the gene bank as described above revealed no significant homologies to MO-39. MO-12, MO-13 and MO-19 were found to show some homologies to unknown sequences previously isolated from *M. tuberculosis*.

EXAMPLE 6

ISOLATION OF DNA SEQUENCES ENCODING M. TUBERCULOSIS ANTIGENS
BY SCREENING OF A NOVEL EXPRESSION LIBRARY

This example illustrates isolation of DNA sequences encoding *M. tuberculosis* antigens by screening of a novel expression library with sera from *M. tuberculosis*-infected patients that were shown to be unreactive with a panel of the recombinant *M. tuberculosis* antigens TbRall, TbRa3, Tb38-1, TbH4, TbF and 38 kD.

Genomic DNA from *M. tuberculosis* Erdman strain was randomly sheared to an average size of 2 kb, and blunt ended with Klenow polymerase, followed by the addition of EcoRI adaptors. The insert was subsequently ligated into the Screen phage vector (Novagen, Madison, WI) and packaged *in vitro* using the PhageMaker extract (Novagen). The resulting library was screened with sera from several *M. tuberculosis* donors that had been shown to be negative on a panel of previously identified *M. tuberculosis* antigens as described above in Example 3B.

A total of 22 different clones were isolated. By comparison, screening of the λ Zap library described above using the same sera did not result in any positive hits. One of the clones was found to represent TbRa11, described above. The determined 5' cDNA sequences for 19 of the remaining 21 clones (hereinafter referred to as Erdsn1. Erdsn2. Erdsn4-Erdsn10, Erdsn12-18, Erdsn21-Erdsn23 and Erdsn25) are provided in SEQ ID NO: 303-322, respectively, with the determined 3' cDNA sequences for Erdsn1. Erdsn2. Erdsn4. Erdsn-5. Erdsn-7-Erdsn10. Erdsn12-18, Erdsn21-Erdsn23 and Erdsn25 being provided in SEQ ID NO: 323-341, respectively. The complete cDNA insert sequence for the clone Erdsn24 is provided in SEQ ID NO: 342. Comparison of the determined cDNA sequences with those in the gene bank revealed no significant homologies to the sequences provided in SEQ ID NO: 309, 316, 318-320, 322, 324, 328, 329, 333, 335, 337, 339 and 341. The sequences of SEQ ID NO: 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342 were found to show some homology to unknown sequences previously identified in *M. tuberculosis*.

EXAMPLE 7 ISOLATION OF SOLUBLE M. TUBERCULOSIS ANTIGENS USING MASS SPECTROMETRY

This example illustrates the use of mass spectrometry to identify soluble M. tuberculosis antigens.

In a first approach, *M. tuberculosis* culture filtrate was screened by Western analysis using serum from a tuberculosis-infected individual. The reactive bands were excised from a silver stained gel and the amino acid sequences determined by mass spectrometry. The determined amino acid sequence for one of the isolated antigens is provided in SEQ ID NO: 343. Comparison of this sequence with those in the gene bank revealed homology to the 85b precursor antigen previously identified in *M. tuberculosis*.

In a second approach, the high molecular weight region of *M. tuberculosis* culture supernatant was studied. This area may contain immunodominant antigens which may be useful in the diagnosis of *M. tuberculosis* infection. Two known monoclonal antibodies, IT42 and IT57 (available from the Center for Disease Control, Atlanta, GA), show reactivity by Western analysis to antigens in this vicinity. although the identity of the antigens remains unknown. In addition, unknown high-molecular weight proteins have been described as containing a surrogate marker for *M. tuberculosis* infection in HIV-positive individuals (*Jnl. Infect. Dis.*, 176:133-143, 1997). To determine the identity of these antigens, two-dimensional gel electrophoresis and two-dimensional Western analysis were performed using the antibodies IT57 and IT42. Five protein spots in the high molecular weight region were identified, individually excised, enzymatically digested and subjected to mass spectrometric analysis.

The determined amino acid sequences for three of these spots (referred to as spots 1. 2 and 4) are provided in SEQ ID NO: 344, 345-346 and 347, respectively. Comparison of these sequences with those in the gene bank revealed that spot 1 is the previously identified PcK-1, a phosphoenolpyruvate kinase. The two sequences isolated from spot 2 were determined to be from two DNAks, previously identified in *M. tuberculosis* as heat shock proteins. Spot 4 was determined to be the previously identified *M. tuberculosis* protein Kat G. To the best of the inventors' knowledge, neither PcK-1 nor the two DNAks have previously been shown to have utility in the diagnosis of *M. tuberculosis* infection.

EXAMPLE 8

Use of Representative Antigens for Diagnosis OF Tuberculosis

This example illustrates the effectiveness of several representative polypeptides in skin tests for the diagnosis of *M. tuberculosis* infection.

Individuals were injected intradermally with 100 µl of either PBS or PBS plus Tween 20TM containing either 0.1 µg of protein (for TbH-9 and TbRa35) or 1.0 µg of protein (for TbRa38-1). Induration was measured between 5-7 days after injection, with a response of 5 mm or greater being considered positive. Of the 20 individuals tested. 2 were PPD negative and 18 were PPD positive. Of the PPD positive individuals, 3 had active tuberculosis. 3 had been previously infected with tuberculosis and 9 were healthy. In a second study, 13 PPD positive individuals were tested with 0.1 µg TbRa11 in either PBS or PBS plus Tween 20TM as described above. The results of both studies are shown in Table 10.

TABLE 10
RESULTS OF DTH TESTING WITH REPRESENTATIVE ANTIGENS

	TbH-9 Pos/Total	Tb38-1 Pos/Total	TbRa35 Pos/Total	Cumulative Pos/Total	TbRa11 Pos/Total
PPD negative	0/2	0/2			
	10/2	0/2	0/2	0/2	
PPD positive					
healthy	5/9	1/9	1/9	6/0	
prior TB	3/5	2/5		6/9	1/4
active	3/4		2/5	4/5	3/5
TOTAL		3/4	0/4	4/4	1/4
TOTAL	11/18	9/18	6/18	14/18	5/13

EXAMPLE 9 SYNTHESIS OF SYNTHETIC POLYPEPTIDES

Polypeptides may be synthesized on a Millipore 9050 peptide synthesizer using FMOC chemistry with HPTU (O-Benzotriazole-N.N.N'.N'-tetramethyluronium

hexafluorophosphate) activation. A Gly-Cys-Gly sequence may be attached to the amino terminus of the peptide to provide a method of conjugation or labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides may be precipitated in cold methyl-t-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0%-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides. Following lyophilization of the pure fractions, the peptides may be characterized using electrospray mass spectrometry and by amino acid analysis.

EXAMPLE 10

PREPARATION AND CHARACTERIZATION OF M. TUBERCULOSIS FUSION PROTEINS

A fusion protein containing TbRa3, the 38 kD antigen and Tb38-1 was prepared as follows.

Each of the DNA constructs TbRa3. 38 kD and Tb38-1 were modified by PCR in order to facilitate their fusion and the subsequent expression of the fusion protein TbRa3-38 kD-Tb38-1. TbRa3, 38 kD and Tb38-1 DNA was used to perform PCR using the primers PDM-64 and PDM-65 (SEQ ID NO: 146 and 147), PDM-57 and PDM-58 (SEQ ID NO: 148 and 149), and PDM-69 and PDM-60 (SEQ ID NO: 150 and 151), respectively. In each case, the DNA amplification was performed using 10 μl 10X Pfu buffer, 2 μl 10 mM dNTPs, 2 μl each of the PCR primers at 10 μM concentration, 81.5 μl water, 1.5 μl Pfu DNA polymerase (Stratagene, La Jolla, CA) and 1 μl DNA at either 70 ng/μl (for TbRa3) or 50 ng/μl (for 38 kD and Tb38-1). For TbRa3, denaturation at 94°C was performed for 2 min, followed by 40 cycles of 96°C for 15 sec and 72°C for 1 min, and lastly by 72°C for 4 min. For 38 kD, denaturation at 96°C was performed for 2 min, followed by 40 cycles of 96°C for 30 sec, 68°C for 15 sec and 72°C for 3 min, and finally by 72°C for 4 min. For Tb38-1 denaturation at 94°C for 2 min was followed by 10 cycles of 96°C for 15 sec and 72°C for

1.5 min, 30 cycles of 96°C for 15 sec, 64°C for 15 sec and 72°C for 1.5, and finally by 72°C for 4 min.

The TbRa3 PCR fragment was digested with NdeI and EcoRI and cloned directly into pT7^L2 IL 1 vector using NdeI and EcoRI sites. The 38 kD PCR fragment was digested with Sse8387I, treated with T4 DNA polymerase to make blunt ends and then digested with EcoRI for direct cloning into the pT7^L2Ra3-1 vector which was digested with StuI and EcoRI. The 38-1 PCR fragment was digested with Eco47III and EcoRI and directly subcloned into pT7^L2Ra3/38kD-17 digested with the same enzymes. The whole fusion was then transferred to pET28b – using NdeI and EcoRI sites. The fusion construct was confirmed by DNA sequencing.

The expression construct was transformed into BLR pLys S *E. coli* (Novagen. Madison, WI) and grown overnight in LB broth with kanamycin (30 µg/ml) and chloramphenicol (34 µg/ml). This culture (12 ml) was used to inoculate 500 ml 2XYT with the same antibiotics and the culture was induced with IPTG at an OD560 of 0.44 to a final concentration of 1.2 mM. Four hours post-induction, the bacteria were harvested and sonicated in 20 mM Tris (8.0), 100 mM NaCl, 0.1% DOC, 20 µg/ml Leupeptin, 20 mM PMSF followed by centrifugation at 26.000 X g. The resulting pellet was resuspended in 8 M urea, 20 mM Tris (8.0). 100 mM NaCl and bound to Pro-bond nickel resin (Invitrogen. Carlsbad, CA). The column was washed several times with the above buffer then eluted with an imidazole gradient (50 mM, 100 mM, 500 mM imidazole was added to 8 M urea, 20 mM Tris (8.0), 100 mM NaCl). The eluates containing the protein of interest were then dialyzed against 10 mM Tris (8.0).

The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbRa3-38 kD-Tb38-1) are provided in SEQ ID NO: 152 and 153, respectively.

A fusion protein containing the two antigens TbH-9 and Tb38-1 (hereinafter referred to as TbH9-Tb38-1) without a hinge sequence, was prepared using a similar procedure to that described above. The DNA sequence for the TbH9-Tb38-1 fusion protein is provided in SEQ ID NO: 156.

The ability of the fusion protein TbH9-Tb38-1 to induce T cell proliferation and IFN- γ production in PBMC preparations was examined using the protocol described above in Example 1. PBMC from three donors were employed: one who had been previously shown to respond to TbH9 but not Tb38-1 (donor 131); one who had been shown to respond to Tb38-1 but not TbH9 (donor 184); and one who had been shown to respond to both antigens (donor 201). The results of these studies (Figs. 5-7, respectively) demonstrate the functional activity of both the antigens in the fusion protein.

A fusion protein containing TbRa3, the antigen 38kD, Tb38-1 and DPEP was prepared as follows.

Each of the DNA constructs TbRa3. 38 kD and Tb38-1 were modified by PCR and cloned into vectors essentially as described above, with the primers PDM-69 (SEQ ID NO:150 and PDM-83 (SEQ ID NO: 205) being used for amplification of the Tb38-1A fragment. Tb38-1A differs from Tb38-1 by a DraI site at the 3' end of the coding region that keeps the final amino acid intact while creating a blunt restriction site that is in frame. The TbRa3/38kD/Tb38-1A fusion was then transferred to pET28b using NdeI and EcoR1 sites.

DPEP DNA was used to perform PCR using the primers PDM-84 and PDM-85 (SEQ ID NO: 206 and 207, respectively) and 1 µl DNA at 50 ng/µl. Denaturation at 94 °C was performed for 2 min. followed by 10 cycles of 96 °C for 15 sec. 68 °C for 15 sec and 72 °C for 1.5 min; 30 cycles of 96 °C for 15 sec. 64 °C for 15 sec and 72 °C for 1.5 min; and finally by 72 °C for 4 min. The DPEP PCR fragment was digested with EcoRI and Eco72I and clones directly into the pET28Ra3/38kD/38-1A construct which was digested with DraI and EcoRI. The fusion construct was confirmed to be correct by DNA sequencing. Recombinant protein was prepared as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-2) are provided in SEQ ID NO: 208 and 209, respectively.

The reactivity of the fusion protein TbF-2 with sera from *M. tuberculosis*-infected patients was examined by ELISA using the protocol described above. The results of these studies (Table 11) demonstrate that all four antigens function independently in the fusion protein.

TABLE 11

REACTIVITY OF TBF-2 FUSION RECOMBINANT WITH TB AND NORMAL SERA

Serum ID	Status	TbF OD450	Status	TbF-2 OD450	Status		ELISA	Reactivity	
			1	1	+	201.0	1		
B931-40	TB	0.57	+	0.321	╅╤──	38 kD	TbRa3	Тъ38-1	DPEP
B931-41	TB	0.601	+	0.396	+	 	+ - -	-	+
B931-109	TB	0.494	+	0.404	-	+	+	+	•
B931-132	TB	1.502	-	1.292	+		+	±	-
5004	TB	1.806	+	1.666	+=	+	+	+	±
15004	TB	2.862	-	2.468	+	±	<u> </u>	+	-
39004	TB	2.443	-	1.722	-	+	+	+	-
68004	TB	2.871	-	2.575	-	 -	<u> </u>	-	-
99004	TB	0.691	-	0.971	 		-	-	
107004	TB	0.875	+	0.732	 	<u> </u>	<u> </u>	! -	-
92004	TB	1.632		1.394		 -	<u> ± </u>	-	-
97004	TB	1.491	-	1.979		+	±	Ξ	
118004	TB	3.182	-	3.045	+	+	±	-	-
173004	ТВ	3.644	_	3.578		+	±	-	-
175004	TB	3.332		2.916		+	-	+	-
274004	TB	3.696		3.716	<u> </u>	 -	+	-	•
276004	TB	3.243			-	<u> </u>	+	-	-
282004	TB	1.249		2.56		<u> • </u>	<u>-</u>	-	
289004	TB	1.373		1.234	-	ļ -	-	-	
308004	ТВ	3.708		1.17		ļ ·	+	-	-
314004	I TB	1 4 6 6		3.355		<u> </u>	•	+	
317004	TB			1.399			-	-	-
312004	TB	700		0.92		<u> </u>	-	-	-
80004	TB	0.238		1.453		<u> </u>	-	-	-
151004	TB		-			<u> - </u>	=	-	_
78004	TB	0.188				-	-	-	=
10004	TB	0.384		0.469	-	•	-	-	Ξ
11004	TB	0.306		2.392	_	Ξ	-	-	
21004	TB	i		0.874		-	-	-	-
28004	TB	0.047		1.456	-	-	-	-	-
6-87	Normai	0.094				•	-		
6-88	Normal	0.21			-	•	-	-	-
6-89	Normal	0.214 -			-		-	-	
6-90	Normal	0.170		0.306	-	-		-	
6-91	Normai	0.179 -				•	-	-	•
6-92	Normal	0.0(1			-	-	-	-	-
	Normal			0.097			-	-	•
	Normal	0.000			-	-	-	-	•
	Normal					-	-		
7.3.1	Normal	0.125		0.159					
	···oiiiai	0.121 -		0.12 .			-		
ut-off		0.284							
0.,		0.284	10	0.266					

A fusion protein containing TbRa3, the antigen 38kD, Tb38-1 and TbH4 was prepared as follows.

Genomic *M. tuberculosis* DNA was used to PCR full-length TbH4 (FL TbH4) with the primers PDM-157 and PDM-160 (SEQ ID NO: 348 and 349, respectively) and 2 µl DNA at 100 ng/µl. Denaturation at 96 °C was performed for 2 min, followed by 40 cycles of 96 °C for 30 sec, 61 °C for 20 sec and 72 °C for 5 min; and finally by annealing at 72 °C for 10 min. The FL TbH4 PCR fragment was digested with EcoRI and Sca I (New England Biolabs.) and cloned directly into the pET28Ra3/38kD/38-1A construct described above which was digested with DraI and EcoRI. The fusion construct was confirmed to be correct by DNA sequencing. Recombinant protein was prepared as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-6) are provided in SEQ ID NO: 350 and 351, respectively.

A fusion protein containing the antigen 38kD and DPEP separated by a linker was prepared as follows.

38 kD DNA was used to perform PCR using the primers PDM-176 and PDM-175 (SEQ ID NO: 352 and 353, respectively), and 1 μ l PET28Ra3/38kD/38-1/Ra2A-12 DNA at 110 ng/μl. Denaturation at 96 °C was performed for 2 min, followed by 40 cycles of 96 °C for 30 sec, 71 °C for 15 sec and 72 °C for 5 min and 40 sec; and finally by annealing at 72 °C for 4 min. The two sets of primers PDM-171, PDM-172, and PDM-173, PDM-174 were annealed by heating to 95 °C for 2 min and then ramping down to 25 °C slowly at 0.1 °C/sec. DPEP DNA was used to perform PCR as described above. The 38 kD fragment was digested with Eco RI (New England Biolabs) and cloned into a modified pT7ΔL2 vector which was cut with Eco 72 I (Promega) and Eco RI. The modified pT7ΔL2 construct was designed to have a MGHHHHHH amino acid coding region in frame just 5° of the Eco 72 I site. The construct was digested with Kpn 2I (Gibco, BRL) and Pst I (New England Biolabs) and the annealed sets of phosphorylated primers (PDM-171, PDM-172 and PDM-173, PDM-174) were cloned in. The DPEP PCR fragment was digested with Eco RI and Eco 72 I and cloned into this second construct which was digested with Eco 47 III (New England Biolabs) and Eco RI. Ligations were done with a ligation kit from Panvera (Madison, WI). The resulting construct was digested with NdeI (New England Biolabs) and Eco RI. and transferred to a

modified pET28 vector. The fusion construct was confirmed to be correct by DNA sequencing.

Recombinant protein was prepared essentially as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-8) are provided in SEQ ID NO: 354 and 355, respectively.

One of skill in the art will appreciate that the order of the individual antigens within the fusion protein may be changed and that comparable activity would be expected provided each of the epitopes is still functionally available. In addition, truncated forms of the proteins containing active epitopes may be used in the construction of fusion proteins.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANTS: Reed, Steven G.
 Skeiky, Yasir A.W.
 Dillon, Davin C.
 Campos-Neto, Antonio
 Houghton, Raymond
 Vedvick, Thomas S.
 Twardzik, Daniel R.
 Lodes, Michael J.
 Hendrickson, Ronald
- (ii) TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS
- (iii) NUMBER OF SEQUENCES: 355
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: SEED and BERRY LLP
 - (B) STREET: 6300 Columbia Center, 701 Fifth Avenue
 - (C) CITY: Seattle
 - (D) STATE: Washington
 - (E) COUNTRY: USA
 - (F) ZIP: 98104-7092
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE: 05-MAY-1998
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Maki, David J.
 - (B) REGISTRATION NUMBER: 31,392
 - (C) REFERENCE/DOCKET NUMBER: 210121.411C9
 - (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (206) 622-4900
 - (B) TELEFAX: (206) 682-6031
- (2) INFORMATION FOR SEQ ID NO:1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 766 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGAGGCACCG	GTAGTTTGAA	CCAAACGCAC	AATCGACGGG	CAAACGAACG	GAAGAACACA	60
ACCATGAAGA	TGGTGAAATC	GATCGCCGCA	GGTCTGACCG	CCGCGGCTGC	AATCGGCGCC	120
GCTGCGGCCG	GTGTGACTTC	GATCATGGCT	GGCGGCCCGG	TCGTATACCA	GATGCAGCCG	180
GTCGTCTTCG	GCGCGCCACT	GCCGTTGGAC	CCGGCATCCG	CCCCTGACGT	CCCGACCGCC	240
GCCCAGTTGA	CCAGCCTGCT	CAACAGCCTC	GCCGATCCCA	ACGTGTCGTT	TGCGAACAAG	300
GGCAGTCTGG	TCGAGGGCGG	CATCGGGGGC	ACCGAGGCGC	GCATCGCCGA	CCACAAGCTG	360
AAGAAGGCCG	CCGAGCACGG	GGATCTGCCG	CTGTCGTTCA	GCGTGACGAA	CATCCAGCCG	420
GCGGCCGCCG	GTTCGGCCAC	CGCCGACGTT	TCCGTCTCGG	GTCCGAAGCT	CTCGTCGCCG	480
GTCACGCAGA	ACGTCACGTT	CGTGAATCAA	GGCGGCTGGA	TGCTGTCACG	CGCATCGGCG	540
ATGGAGTTGC	TGCAGGCCGC	AGGGNAACTG	ATTGGCGGGC	CGGNTTCAGC	CCCCTCTTCA	600
GCTACGCCGC	CCGCCTGGTG	ACGCGTCCAT	GTCGAACACT	CGCGCGTGTA	GCACGGTGCG	660
GTNTGCGCAG	GGNCGCACGC	ACCGCCCGGT	GCAAGCCGTC	CTCGAGATAG	GTGGTGNCTC	720
GNCACCAGNG	ANCACCCCCN	NNTCGNCNNT	TCTCGNTGNT	GNATGA		766

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 752 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATCCATCACC	3.003.003.003					
AIGCAICACC	ATCACCATCA	CGATGAAGTC	ACGGTAGAGA	CGACCTCCGT	CTTCCGCGCA	60
GACTTCCTCA	GCGAGCTGGA	CGCTCCTGCG	CAAGCGGGTA	CGGAGAGGGG	COTOTOCCC	
GTGGAAGGGC	TOCCGCCGG	CTCCCCCC	OTTOCOCCIA	COGRAGAGEGE	GGICICCGGG	120
TOCOCO	72.5555	CICGGCGIIG	CTGGTAGTCA	AACGAGGCCC	CAACGCCGGG	180
TCCCGGTTCC	TACTCGACCA	AGCCATCACG	TCGGCTGGTC	GGCATCCCGA	CAGCGACATA	240
TTTCTCGACG	ACGTGACCGT	GAGCCGTCGC	CATGCTGAAT	TCCCCTTTCC	11101100111	
ייייר א אינייירים. ייייר א אינייירים	TCC3 MCTCCC	22.22	CAIGCIGAAI	TCCGG11GGA	AAACAACGAA	300
TTCAATGTCG	- CGATGTCGG	GAGTCTCAAC	GGCACCTACG	TCAACCGCGA	GCCCGTGGAT	360
TCGGCGGTGC	TGGCGAACGG	CGACGAGGTC	CAGATCGGCA	ACCTCCCCTT		
ACCGGACCCA	AGCNAGGGGN	CCITCICO		AGCICCGGII	GGIGIICIIG	420
ACCGGACCCA	AGCANGGCGA	GGWIGHCGGG	AGTACCGGGG	GCCCGTGAGC	GCACCCGATA	480
GCCCCGCGCT	GGCCGGGATG	TCGATCGGGG	CGGTCCTCCG	ACCTGCTACG	ע ההכיכי ע ההבההה	540
CCCTGATGTC	Caccamence	3 A C A MMCCC 3 M	######################################	ACCIOCIACO	ACCGGMIIII	540
CCCTGATGTC	CACCALL'ACC	AAGATTCGAT	TCTTGGGAGG	CTTGAGGGTC	NGGGTGACCC	600
CCCCGCGGGC	CTCATTCNGG	GGTNTCGGCN	GGTTTCACCC	CNTACCMACT	GCCMCCCCCM	660
TTGCNAATTC	المسالات المسالات	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	221.22	CIVITAC CIVAC I	accure coom	990
TTGCNAATTC		GCCCNNAAAG	GGACCNTTAN	CTTGCCGCTN	GAAANGGTNA	720
TCCNGGGCCC	NTCCTNGAAN	CCCCNTCCCC	CT			752
						154

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 813 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CATATGCATC	ACCATCACCA	TCACACTTCT	AACCGCCCAG	CGCGTCGGG	GCGTCGAGCA	60
CCACGCGACA	CCGGGCCCGA	TCGATCTGCT	AGCTTGAGTC	TEGTENEGEN	TCGTCGTCAG	100
CAGCGCGATG	CCCTATGTTT	GTCGTCGACT	CAGATATOGG	CCCAAMCCAA	TCTCCCGCCT	120
			CAGALAICGC	GGCAAICCAA	TCTCCCGCCT	180

604

GCGGCCGGC CTCCTCC	
GCGGCCGGCG GTGCTGCAAA CTACTCCCGG AGGAATTTCG ACGTGCGCAT CAAGATCTTC	240
ATGCTGGTCA CGGCTGTCGT TTTGCTCTGT TGTTCGGGTG TGGCCACGGC CGCGCCCAAG	300
ACCTACTGCG AGGAGTTGAA AGGCACCGAT ACCGGCCAGG CGTGCCAGAT TCAAATGTCC	360
GACCCGGCCT ACAACATCAA CATCAGCCTG CCCAGTTACT ACCCCGACCA GAAGTCGCTG	420
GAAAATTACA TCGCCCAGAC GCGCGACAAG TTCCTCAGCG CGGCCACATC GTCCACTCCA	480
CGCGAAGCCC CCTACGAATT GAATATCACC TCGGCCACAT ACCAGTCCGC GATACCGCCG	540
CGTGGTACGC AGGCCGTGGT GCTCAMGGTC TACCACAACG CCGCGGCAC GCACCCAACG	600
ACCACGTACA AGGCCTTCGA TTGGGACCAG GCCTATCGCA AGCCAATCAC CTATGACACG CTGTGGCAGG CTGACACCGA TCCGCTGCCA GTCGTCTTCC CCATTGTTGC AAGGTGAACT GAGCAACGCA GACCCCCACACA ATCCGCTGCCA GTCGTCTTCC CCATTGTTGC AAGGTGAACT	660
GAGCAACGCA GACCGGGACA ACWGGTATCG ATAGCCGCCN AATGCCGGCT TGGAACCCNG	720
TGAAATTATC ACAACTTCGC AGTCACNAAA NAA	780
THE STATE OF THE S	813
(2) INFORMATION FOR SEQ ID NO:4:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 447 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
/wi \ CTOTTOTT	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:	
CGGTATGAAC ACCCCCCCCC	
CGGTATGAAC ACGGCCGCGT CCGATAACTT CCAGCTGTCC CAGGGTGGGC AGGGATTCGC	60
TOTAL CONTROL OF THE	120
AICOGCCIA CCGCCTTCCT CCCCTTCCCT CTTCCTCC	180
Transcara Tarres	240
CGGCGACGTG ATCACCGCGG TCGACGGCGC TCCGATCAAC TCGGCCACCG CGATGGCGGA	300
CGCGCTTAAC GGGCATCATC CCGGTGACGT CATCTCGGTG AACTGGCAAA CCAAGTCGGG	360
CGGCACGCGT ACAGGGAACG TGACATTGGC CGAGGGACCC CCGGCCTGAT TTCGTCGYGG ATACCACCCG CCGGCCGGCC AATTGGA	420
TOTAL COURT AND LOCAL	447
2) INFORMATION FOR SEQ ID NO:5:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 604 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
4 11 -	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:	
CTCCCACTCC COMOGGGGGA	
GTCCCACTGC GGTCGCCGAG TATGTCGCCC AGCAAATGTC TGGCAGCCGC CCAACGGAAT	60
TOTAL CONTROL CONTROL CONTROL CONTROL CONTROL CONTROL CONTROL	120
THE TOTAL CONCERNING COCCORDANCE COCCORDAN	180
TOTAL TOTAL TOTAL CONTROL OF THE CON	240
THE CONTRACT CONTRACT CONTRACT AND CONTRACT CONT	300
THE PROPERTY OF THE PROPERTY O	360
TICAGAIGI ICTICGNOGA AMOTGATON CONCORNOS CONCORNOS	420
TOTAL GOIGICCIAN CCCNNNTCC TCGNCGAMAT CANALACTCC ATTOCK	480
NAAAAGGGTG GANCAGNNNN AANTNGNGGN CCNAANAANC NNNANNGNNG NNAGNTNGNT NNNTTTTNNC ANNNNNNNTG NNGNNGNNGN	540
NNATTTNNC ANNNNNNTG NNGNNGNNCN NNNCAANCNN NTNNNNGNAA NNGGNTTNTT	600
	604

(2) INFORMATION FOR SEQ ID NO:6:

(A) LENGTH: 633 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1362 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGACGACGAC	GGCGCCGGAG	AGCGGGCGCG	AACGGCGATC	GACGCGGCCC	#GGGG3 43 45	
CGGCACCACC	CAGGAGGGAG	TCGAATCATC	7 y y managaric	ACCATATTGA	TGGCCAGAGT	60
CCCCGCCGAG	CCGGCGGCGC	GGTCGCCCAC	CHCHARACT	ACCATATTGA	GCCCGTCGCG	120
CGGCTGCCCG	AGCCGCTCGC	CATCOTCTC	GICTATGCCG	AGGCCCGCCG	CGAGTTCGGC	180
GCGACGTTGC				GACTGCTCAC	CGCCGGCTGG	240
GCCGCCGCCG		GCTGGTGGGC	CAGGTGCCGC	GTGGCCGCAA	GGAAGCCGTC	300
	A COCOCCION	CCTGCGCTGC	CCCMGGMGCC	TOCTOCOLOR	63 665 655	360
CIGIACGCGG	CAGGCCAAAC	CGACACCGCC	GCGGCGATCT	TGGCCGGCACA	AGCACCTGCC	420
	COMMCGCGC	GTATGTGGCG	TGGGCGGCNC	CARCCCCCA	10000	
CCGCCGGCAC	CGTTCGGCCC	GGATGTCGCC	GCCGAATACC	TGGGCACCGC	CCTCCAA	480
	CMCGCCIGGI	CCTGGTGCTG	CTGGACGAAA	CCTTCCTGCC	GGIGCAATTC	540
CGCGCCCAAC	AGCTCATGCG	CCGCGCCGGT	GGACTCCTCT			600
GAGCATCGGC	CGGGCCGCTC	CACCCGCCCC	CTCCTCCTCT	TCGCCCGCAA GAACGCTGCC	GGTGCGCGCG	660
GCATGGGCAA	CACCGTCCGA	CCCCATAGO				720
GACACCGCGC	CGCACCTCCC	GCCCATAGCA	ACCGCGTTCG	CCGCGCTCAG	CCACCACCTG	780
TOGCACOCCO	ACCOLLEGE	GCCACCGACT	CGTCAGGTGG	TCAGGCGGGT	CGTGGGGTCG	840
	AGC CHAILGCC	GATGAGCAGT	CGCTGGACGA	ACCACCACAC	CGCCGAGCTG	900
	TO CALCOCOCC	CAUCCGTCTT	CCCCTCCTCT .	~~~~	CCCGCATCAG	960
GIGACCGACG	ACGACGTCGC			ACACCGATGC		1020
GGCGCCCTGG	CCTGGGCCGC	CTTCACCGCC	GCGCGGCGC3		GATCGGCGCC	
GCCGCCGAGG	GCCAGGTGTC	GCGGCAAAAC	CCGACTGGCT	GAGTGTGCGC	GATCGGCGCC	1080
TAGGGTGTCA	TCGCTGGCCC	GAGGGATCTC	GCCCCCCCC	BAGIGIGCGC		1140
GGAAGCTGCG	CCCACTGGCT	TGCGCCCCAA	CCCCCCCCA	ACGGAGGTGG	CGACACAGGT	1200
		GCCCCCAA	CGCCGTCGTG	GGCGTTCGGT	TGGCCGCACT	1260
	GTCACCGCCG	GCCCTTGGCC	GAAGGTCCAG	CTCAACGTGC	CGTCACCGAA	1320
	GTCACCGGGG	GICACCCTGC	GCGCCCAAGG	AA		1362

⁽²⁾ INFORMATION FOR SEQ ID NO:8:

- (A) LENGTH: 1458 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCGACGACCC CGATATGCC	C CCC3 CCC	_			
GCGACGACCC CGATATGCC GTATCGCTCC CGTTGAGGA	G GGCACCGTAC	G CGAAAGCCGT	CGCCGACGC	CTCGGGCGCG	60
GTATCGCTCC CGTTGAGGA TGGATGACGT GGCCCGTGT	C ALLCAGGACT	GCGTGGAGGC	CCGGCTGGGG	GAAGCCCCCTC	120
	- INCHICATO	ACCGGCAGCG	GCGCGCCGAG	CTGCGGACGG	180
CTAAGGCCTT GCTCGGCGTC	3 CGGGACGAGT	TAAAGCTGAG	CTTGGCGGC	GTGACCCTAC	
TGCGCGAGCG CTATCTGCT		AGGCCGGCC	GGCCGAGTCG	ACCGGCGAGC	240
TGATGGACCG ATCGGCGCGC	- 1131141117272				300
CGAGGCGGTG GGCCGAGCGC	a rrecedentice	TATTACCCSA	COTTOCALARMA		360
TOTAL GALGAACIC	GGCACCGACC	TCCCA COLOR			420
CGATTGAGGA TTCGCTGCAA	TCGATCTTTG	CGACGCTGCT	2 CACCGGCTGT	TTTGTTCTGC	480
	TATGCGTTCL	CCCACCMCCC	10000		540
CCTCCACGGG CGGCACGGCC	AGCGGACCCC	TOTOGOG	ACCCGCCGGG	GATCGGGTGG	600
CGGGTGTGGT CTCCATGGGC	GGTCGCCGGC	TGTCGTTTCT			660
CGCACCCGGA TATCTGTGAT	TTCCTCACCC	GIGGCGCCTG	TATGGCTGTG	CTTGATGTGT	720
ATTTCAACCT ATCGGTTGGT	TTCGTCACCG	CCAAGGCCGA	ATCCCCCAGC	GAGCTCCCGC	780
ATTTCAACCT ATCGGTTGGT TACACCGGCT GGTCAATCCG	CCTACCGACG	CGTTCCTGCG	GGCCGTCGAA	CGCAACGGCC	840
					900
					960
					1020
					1020
	GICGMU I GG	Vicinia Company	221222		
	ATCGATGTCA	CCCCCCCAACCC	CTTCCCCGAA	CTCCCTCACC	1140
	A 1 () () () () () () ()	Carronman	TTTGGCGGAA	CTCCTTCCCC	1200
CACIGGGTAT TCCGTACGAC	AGTGAAGAAG	CCCTCCCCC	3 0 0 0 0		1260
GCATACAGCA GGCGGCGCAC	ACGGCATCGC	CCACCCTCCC	GG2 2		1320
	TTCGCGCGGT	CGGGCCCGAC	CCCCA AGCCGG	GGCGCATTCC	1380
CCGTCGCTCC GACGGGCA		COGCCCCAG	GCGCAACGCA	CAGGTCACCT	1440
					1458

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 862 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGACCGGGAT GGGATCGGCG CCGCGCTGCC CATTGCCGCG CCTACAATCT CGTGGTACAA CTGGGCAATC TGCGCTCGCT ATCAGCCGCC GCCGCCGCC GGGCCGGTAC CCGCTCCGGG CGGAGTCTCC CGCGCAAGGC GGATAATTAT TGATCGCTGA GACAACCCCT CGCCTCGTGC CG	GCCGGTTCCG T	TTCATCCTGA	660 720 780 840 862
---	--------------	------------	---------------------------------

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 622 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1200 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GACCAATACG	GCTCCATTCC	GTTGCCCAAA				
AATGCTATTT	Company of Company	T	TCGTTCCAAG	CAAAATTGGC	GGCCGCGGTG	1080
GGGTCGCAAT	TTGGGCCGTA	TCAGCTATTC	CCCCTTCCTCC	acaniaccai	TCCGCAGGTA	1140
			CGGCTGCTGG	GCCGAGGCGG	TCCGCAGGTA GATGGGCGAG	1200

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1155 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCAAGCAGC:	T GCNCCMccmc					
		CTGTTCGACG	AACTGGGCAT	GCCGAAGACC	AAACGCACCA	
AGACCGGCT						60
CGTTTCTGC	ACATCTGCTC	GCCCACCGCC	3007010011	GIICGACAAG	ACCGGGCATC	120
GGTTGCTCC	AGCGGTGGCG	GCCCACCGCG	ACGICACCCG	GCTCAAGGTC	ACCGTCGACG	180
CCGCGACCG		GCCGACGGCC	GCATCCACAC	CACGTTCAAC	CAGACGAMCG	240
70000110000			CCAACCTCCA	GAACATCCCC	ATCCCCALCC	-
				CCCTTTLCCCG	ATCCGCACCG	300
CGGCCGACTA	CAGCCAGATC	GAGATGCGGA		CGGTTACGCC	GAGTTGATGA	360
TCATCGAGGC	GTTCAACACC	GGGGAGGACC		CCTGTCCGGG	GACGAGGGCC	420
GTGTGCCCAT	CCACCACCAC	GGGGAGGACC	TGTATTCGTT	CGTCGCGTCC	CGGGTGTTCG	480
GGCTGGTTTA	- crrcowgg1C	ACCGGCGAGT	TGCGGCGCCG	GGTCAAGGCG	-	
		GCCTACGGCC				540
AAGCCAACGA	GCAGATGGAC	GCGTATTTCG	CCCCAmmaga	GIIGAAAAIC	TCCACCGAGG	600
GCGCCGTAGT	CGAGCGGGCC	CGCAAGGACG			GACTACCTGC	660
GCTACCTGCC	CGAGCTCCAG	COCAMOGACG	GCTACACCTC	GACGGTGCTG	GGCCGTCGCC	720
CGCTGAACCC	COROCIGGAC	AGCAGCAACC	GTCAAGTGCG	GGAGGCCGCC		
COCTONACEC	GCCGATCCAG	GGCAGCGCCC	CCC2			780
1 CONCLARGGC	GCTCAACGAG	GCACAGCTGG	CCTCCCCCA	CANGGI GGCC		840
AGCTGCTGTT	CGAAATCGCC	CCCGCTCAAG			GTCCACGACG	900
AGATGGGCGG	~~~~	CCCGGTGAAC	GCGAGCGGGT	CGAGGCCCTG	GTGCGCGACA	960
GCTGGGACGC		CTCGACGTCC	CGCTGGAGGT	GTCGGTGGGC	TACGGCCGCA	
	GGCGGCGCAC	TGAGTGCCGA	GCGTGC2TCT	~~~~		1020
TTTCCGCCCT	GAGTTCACGC	TCGGCGCAAT			TTCGGCGATT	1080
CGAGTAGCCT	CGTCA		COCOACCOAG	TTTGTCCAGC	GTGTACCCGT	1140
						1155

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1771 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ATCGAAGACA GGCCGTCGGA ATCGAGAACT ACCTAGTTGT GGCCCGAGTA GTATTCGCCA GTACAGCCAG	CCGACCCAGG TGCCGATTCC CTCGGGGTTC GCAGTTACTG GTGGGCCTAG CCGCCGCAGC CAGTTCGACT	CGGCCAGCCC GCAGCTTCCC	CCCAGGTTAT	ACCACGGCGG CGATTTACGT CTGGCGCTCG AATCTCAGTC AGTCCACGCA GACATGACGA GCTCAGGGGC	AGCAGCAAAC	60 120 180 240 300 360 420 480 540
--	--	--------------------------	------------	--	------------	--

		TGGTGTCCTT		CGTGCAGGCA GGCGCGGCCG		660 720
AAGCATCCC GCCCAGTGTC	CONTACA		GTCGGTCGAA	GGCGCGGCCG GTGGCTGCCA CAGGTGGCGG TCGGAGGAGG	GCGCGGCGCC	780 840
	2000000	CGCCGAAAAC	CAACAACCAC GACGGTAACC	GTGATCGCGG TTCTCTGACG	CGGCCGCCAA GGCGGACCGC	900 960 1020
CGTCTCCGGG GGTGCTGGCG	CTCACCCGA ATCGGGTCGC	TCTCCCTGGG CGCTCGGTTT	CAGTGATATC TTCCTCCTCG GGAGGGCACC	GCCGTCGTCC GACCTGAGGG GTGACCACGG	GTGTTCAGGG TCGGTCAGCC	1080
CGCCATTCAG GAACGCTCAA	CTCGTCGGAG	CGACGACCGG CGATCAACCC TCAACTCGGC	CGAGGCCGGC	AACCAGAACA GGGGGCGCGC	CCGTGCTGGA TGGTGAACAT	1200 1260 1320
TOCGCAGAGC	GGCTCGATCG TTGATCAGCA	GTCTCGGTTT CCGGCAAGGC	TGCGATTCCA GTCACATGCC	GTCGACCAGG TCCCTGGGTG	ACTCAGCCGA CCAAGCGCAT TGCAGGTGAC	1380 1440 1500
GAACGCTGGA CGCGGACGCG	GTGCCGAAGG TTGGTTGCCG	~~~~	CGTCGAAGTA CACCAAGGTC CAAAGCGCCG	GACGACCGCC	GTGCTGCCGC CGATCAACAG TGGCGCTAAC	1560 1620
CTTTCAGGAT GTGATGAAGG	CCCLCGGGCG	GTAGCCGCAC	ACTICANOTA	ACCCTCGGCA	AGGCGGAGCA	1680 1740 1771

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1058 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTCCACCGC	GTGGCGGCCG	CTCTAGAACT	' AGTGGATCCC	CCGGGCTGCA		
ACGAGGATCO	GACGTCGCAG	Cambancay y C	AGIGGAICCC	CCGGGCTGCA	GGAATTCGGC	60
AGCCCGGCGA	CGGCGAGCGC	CCCTTCCAMC	CCGCCGCCGC	GGAAGTATCG	GTCCATGCCT	120
CCGGCGACGG		CGGAATGGCG	CGAGTGAGGA	GGCGGGCAAT	TTGGCGGGGC	180
ATCCAATCAA		AATGGCGCGA	GTGAGGAGGC	GGGCAGTCAT	GCCCAGCGTG	240
	CCTGCATTCG	GCCTGCGGGC	CCATTTGACA	ATCGAGGTAC	TGAGCGCAAA	
TGAATGATGG	AAAACGGGCG	GTGACGTCCG	مدر المساملين المارين	CC#CC#1		300
CGTTGTGGCT	ATCAGGATGT	TCTTCGCCGA	AACCTCATCC		GCCTGCCTGG	360
TGAGCCCGAC	GGCGTCCGAC	CCCCCCCTCC	ARCCIGATGC		GGTGTTCCCG	420
CAAAAGGGTT	GACCAGCGTG				CTTGATGCGA	480
TGGGTATTAC	CACTCCCCC	CACGTAGCGG	TCCGAACAAC	CGGGAAAGTC	GACAGCTTGC	540
CCA CCTA CA 3	CAGTGCCGAT	GTCGACGTCC	GGGCCAATCC	GCTCGCGGCA	AAGGGCGTAT	600
		GGIGICICI	THE PROPERTY OF THE PROPERTY O	10000	ATCTCCCTCA	
		AATCTCGGCT	CGATTTCTCA	ACTOTORACIO		660
TCGATCCTGC	CGCTGGGGTG	ACGCAGCTGC	TCTCCCCTCT		TCACGCGTGC	720
GTACCGAAGT	GATAGACGGA	ATTTCCACCA	TGTCCGGTGT	CACGAACCTC	CAAGCGCAAG	780
CTGTCAAGAT	GCTTGATCCT	CCCCCA	CCAAAATCAC	CGGGACCATC	CCCGCGAGCT	840
			GTGCAAGGCC	GGCGACCGTG	TGGATTGCCC	900
		GTCCGAGCGA	GCATCGACCT	CGGATCCGGG	TCGATTCAGC	960
	GAAATGGAAC	GAACCCGTCA	ACCTCCACTA	annes		
GTTGNTCGAA	ACGCCCTTGT	GAACGGTGTC	AACGGMAC	JUCCOARGII	GCGTCGACGC	1020
			. a se coulde			1058

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 542 base pairs
 - (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAATTCGGCA GGCGGCGGAG CATTGAGCAG AGTGTCGTTC AAATCGCACG GCGCGGCCCA CCGGAGTTAA CCGTGGCCAG AGCGTCCGTA GG	GACATGGCCG AAGATGAGGC GTTTGCGGTT GGTCCGCGTG TGCTTCGCGT CCCGTCGATG	TGGACAGCGC CGGCGCAACC GATTCGTGCG CTGCCGTATC CGACCCGAAC	TAGCGTCGAT CGGCAAGATC GCGCTAGCAC ATTTTGTGTC CAGGCGTGCA TGGGCGATCC	GACATCCGCG ACCTACCGCA GGGCCGGCGA TGCTCGCCGA TCGCGATTCC GCCGGNGAGC	TCAAGCTCGA GCAAGACGCA GGCCTACCAG GGCGGCCACG TGATCGATGA	60 120 180 240 300 360 420 480 540
--	--	--	---	--	--	--

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 913 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CGGTGCCGCC	CGCGCCTCCG	TTGCCCCCAT	TGCCGCCCTC	CCCC1 mc1	TGCGCATCGC	
CACCATCACC	GCCTTTGCCG	CCGCCACCCC	CCCCCCCCCCC	GCCGATCAGC	TGCGCATCGC	60
TTGACCCTGG			CGGTGGCGCC	GGGGCCGCCG	ATGCCACCGC	120
GCCGTCGCCA	CCGTCGCCGC	CGCCATTGCC	ATACAGCACC	CCGCCGGGGG	CACCGTTACC	180
CAAGCCCGCC	GCCCCCT CCC	CGCTGCCGTT	TCAGGCCGGG	GAGGCCGAAT	GAACCGCCGC	240
CCGAACACCC	GCCGGCACCG	TIGCCGCC	TITCCCCCCCC			300
	· a rockice of I	accidence acrem		M1100000		360
	accept TMC		Carried			
		GGGCACCCC	363666666	77777 A		420
		GGTGCCGCCG	GCCCCGCCGT	TTCCCCCCA	73.555555	480
	- recoccot	AATGTTTATC	AACCCCCCCC		CACCGGCCAT	540
CCGGGCGCCG	GAGNGCGTGC	CCGCCGGCGC	CCCCARGO	CGCCAGCGCG	GCCCCTATTG	600
CGGCCCCGCC	GGACCCACCG	GTCCCCCCC	CGCCAACGCC	CAAAAGCCCG	GGGTTGCCAC	660
TGGTGCTGCT	GAAGCCCCTTTA	CCCCCCCCGA	TCCCCCCGTT	GCCGCCGGTG	CCGCCGCCAT	720
	- LIGCEGIIA	GCGCCGGTTC	CGCSGGTTCC	GGCGGTGGCG	CCNTGGCCGC	780
	CCCCCCTAC	AGCCACCCC	CGGTGGCGCC	GTTGCCGCCA	TTGCCGCCAT	840
·	GCCGCCATTG	CCGCCGTTCC	CGCCGCCACC	GCCGGNTTGG	CCGCCGGCGC	900
CGCCGGCGGC	CGC					-
						913

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1872 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GACTACGTTG GTGTAG	AAAA ATCCTGCCCC	CCCC3 222			
TAGCTACCCC GACACA	GGAG GTTAGGGGG	CCGGACCCTT	AAGGCTGGGA	CAATTTCTGA	60
TAGCTACCCC GACACA	TADDEJALIE CACE	GAGCAATTCG	CGCCGCCGCT	CACTCAGGTG	120

GTCATGGTTG CTGAGCGTG GGCGGCCCCG CCGCCTTG	C TGGCTGCCG	T CGGGCTTCCC			
446666110	1 CUCAGGACTY	2 CAMPOOUTON A	* ************************************		180.
	C WWG TOGGGGGG	, PLACEMOCHIC			240
	G CCGGGACTT	: ^3\m^~m~~			300
areardy (יאור ארטייט לארט ליבור א			360
CCAAACCTAC GGCGTCGATC	G TGGTCGGGT2	TGACCCCAAL	GCGTTCAGCG	TCGGCTCCGG	420
	- 101111111	: CCCCIMAAAA	~~~~~~		480
	A ALAGCIAGO		3.0000		540
CAGGGTGGTC GCGCTCGGCC	AAACCGTGCA	CCCCCCCCC	ACGCCCCGTG	CGGTGCCTGG	600
GACATTGAAC GGGTTGATCO	AGTTCGATCC	CCCAATCGAT	TCGCTGACCG	GTGCCGAAGA	660
CGTCGTCAAC GGCCTAGGAC	AGGTGGTCGG	TATCARC	CCCGGTGATT	CGGGCGGGCC	720
GCTGTCCCAG GGTGGGCAGG	GATTCGCCAT	TATGAACACG	GCCGCGTCCG	ATAACTTCCA	780
CCAAATCCGA TCGGGTGGGG	GGTCACCCAC	CCTTCATCGGG	CAGGCGATGG	CGATCGCGGG	840
CTTGGGTGTT GTCGACAACA TCCGGCGGCA AGTCTCGGCA	ACGGCAACCC	CGITCATATC	GGGCCTACCG	CCTTCCTCGG	900
TCCGGCGGCA AGTCTCGGCA GATCAACTCG GCCACCGCGA	TCTCCACCGC	CGCACGAGTC	CAACGCGTGG	TCGGAAGCGC	960
GATCAACTCG GCCACCGCGA	TGGCGGACGC	CGACGTGATC	ACCGCGGTCG	ACGGCGCTCC	1020
CTCGGTGAAC TGGCAAACCA	AGTEGGGGG	GCTTAACGGG	CATCATCCCG	GTGACGTCAT	1080
GGGACCCCCG GCCTGATTTG	AGTCGGGCGG	CACGCGTACA	GGGAACGTGA	CATTGGCCGA	1140
CAGCCGTGAT TGCCGCGTGA	GCCCCCGACT	CACCCGCCGG	CCGGCCAATT	GGATTGGCGC	1200
GCAATGAACG AGGCAGAACA GGCGGTGTGG TCGAGCATCC	GCCCCCGAGT	Chestrates	GTGCGCGTGG	CATTGTGGAA	1260
GGCGGTGTGG TCGAGCATCC	GGATGCCAAC	CACCCTCCCG	TGCAGGGCAG	TTACGTCGAA	1320
GATCCGACCT GGTTTAAGCA GCCAGCGCGG ACGGTTCCGN	CGCCGTCTTC	GACTTCGGCA	GCGCCGCCGC	CCTGCCCGCC	1380
GCCAGCGCGG ACGGTTCCGN TGGCTTGGCAT	CGATCTCCCT	TACGAGGTGC	TGGTCCGGGC	GTTCTTCGAC	1440
TGGCTTGGCA TCGACTGCAT GGTTACGACA TTCGCGACTT	CERTECTOCGI	GGACTCATCG	ATCGCCTCGA	CTACCTGCAG	1500
GGTTACGACA TTCGCGACTT GTCGCCCTGG TCGACACCGC	CTACAACCTC	CULTCCTACG	ACTCACCGCT	GCGCGACGGC	1560
GTCGCCCTGG TCGACACCGC AATCACACCT CGGAGTCGCA	TCACCCCCCA	CIGCCCGAAT	TCGGCACCGT	CGACGATTTC	1620
AATCACACCT CGGAGTCGCA	CCCCCCCCCA	GGTATCCGCA	TCATCACCGA	CCTGGTGATG	1680
TACGGTGACT ATTACGTGTG TTCGTCGACA CCGAAGAGTC	CCCCTGGTTT	AGGAGTCCC	GCCGCGACCC	AGACGGACCG	1740
TTCGTCGACA CCGAAGAGTC	GAACTGGTC	AGCGAGCGCT	ACACCGACGC	CCGGATCATC	1800
TTCGTCGACA CCGAAGAGTC GCACCGATTC TT	OFFICE GICA	LICGATCCTG	TCCGCCGACA	GTTNCTACTG	1860
					1872

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1482 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTTCGCCGAA	ACCTGATGCC	GAGGAACAGG	GTCTTCCCCC	23.22222		
CCGCGCTCCT	CGCCGAGATC	AGGCAGTCCC	OTGITCCCGT	GAGCCCGACG	GCGTCCGACC	60
ACGTAGCGGT	CCGAACAACC	AGGCAGTCGC	TTGATGCGAC	AAAAGGGTTG	ACCAGCGTGC	120
TOGACGTOCS	CCCCIAMORE	GGGAAAGTCG	ACAGCTTGCT	GGGTATTACC	AGTGCCGATG	180
		CIUGUGGGAA	ACCCCCCCANCC	43 44 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4		240
	- COOCIACAA	GGCGACAACA	لا لا ماليات السال	3 000000000		
	CHILLCIOMA	CIGICAACTT		003 0000		300
CGCAGCTGCT	GTCCGGTGTC	ACGAACCTCC	AAGCCCAACC	CGAICCIGCC	GCTGGGGTGA	360
TTTCGACCAC	CAAAATCACC	GGGACCATCC	ARGCGCAAGG	TACCGAAGTG	ATAGACGGAA	420
GCGCCAAGAG	TGCAAGGCCC	CCCACCAICC	CCGCGAGCTC	TGTCAAGATG	CTTGATCCTG	480
TCCGAGCGAG	CATCCACCTC	GCGACCGTGT	GGATTGCCCA	GGACGGCTCG	CACCACCTCG	540
		GGATCCGGGT		C1 CCC1		600
	COMC IMG	GCCGAAGTTC	CCTCCACCAC	THE		
	***************************************	AMMAC TGACE	CCCTGACGCC	ATCTCAAAAT	TCLCCCTTGTG	660
	COLUGITA	TACTICEGING		CMCCCC CCC		720
CGGTCTTTGA	GCCGGTAGCT	GLCGCCccccc	3000000000	GTGGGACGCG	GCCGAGGTCG	780
			AGGGGGACGA	CTTCAGCATG	GTGGACGAGG	840

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 876 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CGAGCCGGCG	ATAGCTTCTG	GECCCCCCC	G1 001 00		
GCCACCGCCG	GCCCCACCAC	00000000000	GACCAGATGG	CTCGAGGGTT	60
THE CHECKER	3000CACCAC	CCTGACCGGT	GAGGGCCTGC	AACACGCCGA	120
	ALGUATTAA		~~~~		180
	T C CONTRACTOR AND A CO.				
	T CALCUTATION A		73.00me		240
GATCCCGAGG	GCGTGCTGGG	CCCTATOR	TACGIGCAGC	CGCCGGAGCC	300
AACAAGGNGC	ACATCOTTCC	GGGIAICIAC	CGNTATCACG	CGGCCACCGA	360
ATCCTCCCC	AGAICCIGGC	CTCCGGGGTA	GCGATGCCCG	CGGCGCTGCG	420
	CCGAGTGGGG	TOTO COCO	43 aaaaaa		480
	AND DESCRIPTION OF THE PARTY OF		~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~		
	ACCIGACTAG	366666666	3300000		540
GACTGGATGC	GCGCGGTCCC	CCACCACAMO	AATGCTCGGG	GCCCGGTGAT	600
TTGGGCACCG	7000001000	COMGCAGATC	CGACCGTGGG	TGCCGGGCAC	660
100000000	wedgettege	TITTTCCGAC	ACTCGGCCCG	CCGGTCGTCG	720
	AAIUUUAGGT	TECTOCCOOM			780
	CALICGUIGO		CCGCCCGCCC	ACTTACCCC	
GGTGGGGGGT	TGCGCCCGAN	TAACTT		MOTIMUCCOGG	840
					876
	TTGCTGCTGG ATCGGCTACA TTCTTCTACA GATCCCGAGG AACAAGGNGC ATGCTGGCCG CTAAACCGCG GGCGTGCCCT GACTGGATGC TTGGGCACCG ACCGACGCCG AATATCGACC	TTGCTGCTGG ACGCACCAA ATCGGCTACA TCGNGGAAAG TTCTTCTACA TCACCGTCTA GATCCCGAGG GCGTGCTGGG AACAAGGNGC AGATCCTGGC ATGCTGGCCG ACGGGTGGT GGCGTGCCCT ACGTGACGAG GACTGGATGC TTGGGCACCG ACGGGTTCGG ACCGACGCCG AATCCCAGGT AATATCGACC CATTCGGTGC	TTGCTGCTGG ACGCACCAA CCCGGCGGTG ATCGGCTACA TCGNGGAAAG CGGACTGGCC GATCCCGAGG GCGTGCTGG GGGTATCTAC AACAAGGNGC AGATCCTGGC CTCCGGGGTA ATGCTGGCG CCGAGTGGGA TGTCGCCGGC CTAAACCGCG ACGGGTGGT CATCGAGACC GGCGTGCCCT ACGTGACGAG AGCGCTGGAG GACTGGATGC GCGGGTTCCC CGAGCAGATC TTGGGCACCG ACGGGTTCGG TTTTTCCGAC ACCGACGCG AATCCAGGT TGGTCGCTGT	TTGCTGCTGG ACGCCACCAC CCTGACCGGT GAGGGCCTGC ATCGGCTACA TCGNGGAAAG CCGGACTGGC AGGATGTGCG GATCCCGAGG GCGTGCTGG GGGTATCTAC CGNTATCACG AACAAGGNGC AGATCCTGGC CTCCGGGGTA GCGATGCCCG ATGCTGGCCG CCGAGTGGGA TGTCGCCGCC GACGTGTGGT CTAAACCGCG ACGGGTGGT CATCGAGACC GAGAAGCTCC GGCGTGCCCT ACGTGACGAG AGCGCTGGAG AATGCTCGGG GACTGGATGC GAGCAGATC GACCGGGTCCC CGAGCAGATC CGACCGTGGG ACCGGACGCCG AATCCCAGGT TGGTCGCGGT TTTTTCCGAC ACTCGGCCCG AATATCGACC CATTCGGTGC CCCGCCTGCGAGACCGAGACGCCG AATATCCAAGGT TGGTCGCGGT TTTTTGGGACGG	GCAGCCGCG GGCGCACCAC GCCACCGCG GGCGCACCAC TTGCTGCTGG ACGCCACCAC ATCGGCTACA TCGNGGAAAG CGGCACGGC GGGGTGCC GAGCGTACAA TCCACCGTCTA TCACCGTCTAC ATCCGCACGAG GCGGTCCTAC GAGCATGGCC AGGATGTGCG GGGAGAACCC TTCTTCTACA TCACCGTCTAC GAACCAGGCCG TACGTGCAGC CGCCGGAGCCC ACACAAGGNGC AGATCCTGGC CTCCGGGGTA ACAAAGGNGC ACGCGTGGTAC CTCCGGGGTACCAC ACGTGCCGC CCGAGTGGGA AGCCGTGCGC CTAAACCGGC ACGGGTGGT CATCGAGACCC GACGTGCCCT ACGTGACGAG AGCCGTTGGG GACTGGATGC GACGTGCCCT ACGTGACGAG AGCGCTGGGG GACTGGATGCC GACGAGACCCC TTGGGCACCG ACGGGTTCGC CGAGCAGATC CGACCGTGGG ACCGACGCCG AATCCCAGGT TGGTCGCGGT TTTTTCCGAC ACTCGGCCC AGTTACCCGG GGTGGGGGT TGGCCCGAACCC AATCCCAGGT TGGTCGCGGT TTTTGGGAGGG GTTGGCCGGG AATCCCAGGT TGGTCGCGGT TTTTGGGAGGG GTTGGCCGGG GGTGGGGGGT TGCCCCGAACTC CGGTCGTCGC CCGCCCCCAACTCCCGCCC AGTTACCCGG

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1021 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATCCCCCCGG	GCTGCAGGAA ACGAATTCAC	TTCGGCACGA	GAGACAAAA	TCC3 222		
CAGATTCATA	ACGAATTCAC	AGCGCCACAA	GAGACAMAAI	TCCACGCGTT	AATGCAGGAA	60
AGCGAAGACC	TGCCGCACTT	AGCGGCACAA	CAATATGTCG	CGATCGCGGT	AATGCAGGAA TTATTTCGAC	120
	CCCCACII	Carrier Carrier Carrier	The second secon	* * - ·		180
	- CCICCICCA	ACAC CTTTT	CACCOCCACA			240
						300
GATTTCCTCG	GCGAGCAGTT	CATGCAGTCC	TTCTTCALAG	CGGTGGCCCG	CGACGAGGGC	360
			LICILGCAGG	AACAGATCGA	AGAGGTGGCC	420

TTGATGGCAA CCCTCGTGCG CCTTGGGGG	
TTGATGGCAA CCCTGGTGCG GGTTGCCGAT CGGGCCGGGG CCAACCTGTT CGAGCTAGAG	480
	540
	660
	720
	780
	840
	900
GAGCGCCAGC AGTTGTTTTT CCACCAGCGA AGCGTTTTCG GGTCATCGGN GGCNNTTAAG	1020
	1021
(2) INFORMATION FOR SEQ ID NO:21:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 321 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
101010G1: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:	
CGTGCCGACG AACGGAAGAA CACAACCATG AAGATGGTGA AATCGATCGC CGCAGGTCTG	
	60
	120
	180
	240
GGNGNGNATC GNCGANCACA A	300
(2)	321
(2) INFORMATION FOR SEQ ID NO:22:	
(i) CENTRAL	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 373 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(Xi) SEOMENCE DESCRIPTION	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:	
TCTTATCGGT TCCGGTTGGC GAGGGGTTGG	
TCTTATCGGT TCCGGTTGGC GACGGGTTTT GGGNGCGGGT GGTTAACCCG CTCGGCCAGC	60
CGATCGACGG GCGCGGAGAC GTCGACTCCG ATACTCGGCG CGCGCTGGAG CTCCAGGCGC CCTCGGTGGT GNACCGGCAA GCCGTGAACC ACCCTGAGCGC	120
CCTCGGTGGT GNACCGGCAA GGCGTGAAGG AGCCGTTGNA GACCGGGATC AAGGCGATTG	180
	240
	300
GGTGGATCCC AAGAAGCAGG TGCGCTTGTG TATACGTTGG CCATCGGGCA AGAAGGGGAA CTTACCATCG CCG	360
	373
(2) INFORMATION FOR SEQ ID NO:23:	
0% ODG ID MO:23:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 352 base pairs	
(B) mrbs	

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTGACGCCGT TGGTTACCCG TCTTGACGGC TGATCCATGC GCGTGGAGGT TTGACGACGA	CTGGTACGGG CGGTACCGGC TTTCGTCACC	TTGGCCGATT GGTGTGGGCA GCCAGCCGTG	TTGCTCAGGC TAGCCGAGAT TGGCGGCTGT	AGCCGCTGTG CAAGGCGGGC GCAGCTGGCT	CCGGTGGTGT GAATCGGTGC CGCCAGTGGG	60 120 180 240 300 352
					U.F.	352

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 726 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:24:

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 580 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CTGCCGATG ATAGGGAACA GGCGGGGCAT AGCACTCAAT AGGGGCGCGG TGACACCCGA AGCCCAACGG	GCGGCCCGGT ATAGGGGGGT GCTCGGCGCC GGCGGCGATG CATTGTGATG CGAAGCCGCC CGAATGGTCG	GAAGTCATTG GATTTGGCAG GACCAGGCTC AAGCCCCGGA CGAGTACCAC GCACTGGGTG GCGTTACGCG	GCCCACTGAG CGCCGGGGCT TTCAATGTCG GCGCAGGCGG CCGGCGACGG TTGAGGGTGG ACGAACTCAA	TTGATGCAGG CGACCTTTTG TGTGCACCTG GGTATGGCTG GCCAGCCCGA TCCTTTGGAA CGGTCGCCTG AGGCGTTACT GGTAGATGTC	ACCAGCCGGG ATGAACCCGA GAAATCCAAT ATCTGGAGGG GCAACTAAGG GTCGTCGAGC	60 120 180 240 300 360 420 480 540
TCGGCGATGT	ATGCCCAGGA	GAACTCTTGG	ATACAGCGCT	GGTAGATGTC	CAGTGTCTGC	540 580

⁽²⁾ INFORMATION FOR SEQ ID NO:26:

(A) LENGTH: 160 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:	
AACGGAGGCG CCGGGGGTTT TGGCGGGGCC GGGGCGGTCG GCGGCAACGG CGGGGCCGGC GGTACCGCCG GGTTGTTCGG TGTCGGCGGG GCCGGTGGGG CCGGAGGCAA CGGCATCGCC GGTGTCACGG GTACGTCGGC CAGCACACCG GGTGGATCCG	60 120 160
(2) INFORMATION FOR SEQ ID NO:27:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 272 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:	
GACACCGATA CGATGGTGAT GTACGCCAAC GTTGTCGACA CGCTCGAGGC GTTCACGATC CAGCGCACAC CCGACGGCGT GACCATCGGC GATGCGGCCC CGTTCGCGGA GGCGGCTGCC AAGGCGAAC AGTGGACCA AGTGGACCA AGTGGACCA AGTGGACCA AGTGGACCA CGCCAACACA ACGTTGGCGT TGGCGCCCGG TGTCGTTGTC GCCTACGAGC GCAACGACA GACCAACGCC CG	60 120 180 240 272
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 317 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:	
GCAGCCGGTG GTTCTCGGAC TATCTGCGCA CGGTGACGCA GCGCGACGTG CGCGAGCTGA AGCGGATCGA GCAGACGGAT CGCCTGCCGC GGTTCATGGC CTACCTGGCC GCTATCACCG CGCAGGAGCT GAACGTGGCC GAAGCGGCGC GGGTCATCGC GGTCGACGCG GGGACGATCC GTTCGGATCT GGCGTGGTTC GAGAAGATCCA ACTCGGTACA TCGCCTGGCC GCCTGGTCGC CGGCCTGGTT GCGCGGG	60 120 180 240 300 317
(2) INFORMATION FOR SEQ ID NO:29:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 182 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:	

GATCGTGGAG CTGTCGATGA ACAGCGTTGC CGGACGCGCG GCGGCCAGCA CGTCGGTGTA GCAGCGCCGG ACCACCTCGC CGGTGGGCAG CATGGTGATG ACCACGTCGG CCTCGGCCAC CGCTTCGGGC GCGCTACGAA ACACCGCGAC ACCGTGCGC GCGGCGCCGG ACGCCGCCGT GG	60 120 180 182
(2) INFORMATION FOR SEQ ID NO:30:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 308 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:	
GATCGCGAAG TTTGGTGAGC AGGTGGTCGA CGCGAAAGTC TGGGCGCCTG CGAAGCGGGT CGGCGTTCAC GAGGCGAAGA CACGCCTGTC CGAGCTGCTG CGGCTCGTCT ACGGCGGGCA GAGGTTGAGA TTGCCCGCCG CGGCGAGCCG GTAGCAAAGC TTGTGCCGCT GCATCCTCAT GAGACTCGAC GGTTAGGCAT TGACCATGGC GTGTACCGCG TGCCCGACGA TTTGGACGCT CCGTTGTCAG ACGACGTGCT CGAACGCTTT CACCGGTGAA GCGCTACCTC ATCGACACCC	60 120 180 240 300 308
(2) INFORMATION FOR SEQ ID NO:31: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 267 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:	
CCGACGACGA GCAACTCACG TGGATGATGG TCGGCAGCGG CATTGAGGAC GGAGAGAATC CGGCCGAAGC TGCCGCGGG CAAGTGCTCA TAGTGACCGG CCGTAGAGGG CTCCCCCGAT GGCACCGGAC TATTCTGGTG TGCCGCTGGC CGGTAAGAGC GGGTAAAAGA ATGTGAGGGG ACACGATGAG CAATCACACC TACCGAGTGA TCGAGATCGT CGGGACCTCG CCCGACGGCG TCGACGCGGC AATCCAGGGC GGTCTGG	60 120 180 240 267
(2) INFORMATION FOR SEQ ID NO:32:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1539 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:	
CTCGTGCCGA AAGAATGTGA GGGGACACGA TGAGCAATCA CACCTACCGA GTGATCGAGA TCGTCGGGAC CTCGCCCGAC GGCGTCGACG CGGCAATCCA GGGCGGTCTG GCCCGAGCTG CGCAGACCAT GCGCGCCGTG GACTGGTTCG AAGTACAGTC AATTCGAGGC CACCTGGTCG ACGGAGCGGT CAGGTGACTA TGAAAGTCGG CTTCCGCTGG AGGATTCCTG AACCTTCAAG CGCGGCCGAT AACTGAGGTG CATCATTAAG CGACTTTTCC AGAACATCCT GACGGGCTCG AAACGCGGTT CAGCCGACGG TGGCTCCGCC GAGGCGCTGC CTCCAAAATC CCTGCGACAA TTCGTCGGCG GCGCCTACAA GGAAGTCGGT GCTGAATTCG TCGGGTATCT	60 120 180 240 300 360 420

GGTCGACCTG						
	-010000100	AGCCGGACGA	AGCGGTGCTC	GACGTCGGCT	GCGGCTCGGG	480
GCGGATGGCG	TIGCCGCTCA	CCGGCTATCT	GAACAGCGAG	GGACGCTACG	CCGGCTTCGA	
TATCTCGCAG	AAAGCCATCG			ACCTCGGCGC		540
CCAGTTCGAG	GTCTCCGACA	TCTACAACTC	CCTCTACAAC			600
ACTAGACTTT	CGCTTTCCAT	ATCCGGATGC	GETGIACAAC		AATACCAGTC	660
GTTCACCCAC	ATCTTTCCCC	ATCCGGATGC	GTCGTTCGAT	GTGGTGTTTC	TTACCTCGGT	720
GNACCCCCC	AIGITICCGC	CGGACGTGGA	GCACTATCTG	GACGAGATCT	CCCGCGTGCT	780
CARGCCCGGC	GGACGATGCC	TGTGCACGTA	CTTCTTGCTC	AATGACGAGT	CGTTAGCCCA	840
CATCGCGGAA	GGAAAGAGTG	CGCACAACTT	CCAGCATGAG	GGACCGGGTT	ATCGGACAAT	900
CCACAAGAAG	CGGCCCGAAG	AAGCAATCGG	CTTGCCGGAG	ACCUTCATO	CCCATCTCT	
TGGCAAGTTC	GGCCTCGCCG	TGCACGAACC	ATTGCACTAC	GGCTCATGGA	GGGATGTCTA	960
ACCACGCCTA	AGCTTCCAGG	ACATCCTCAT	CCCCTACIAC	GGCTCATGGA	GTGGCCGGGA	1020
		ACATCGTCAT	CGCGACCAAA	ACCGCGAGCT	AGGTCGGCAT	1080
			GCGCCGCTGC	CGGCAGGCCG	ATTAGGCGGG	1140
CAGATIAGCC	CGCCGCGGCT	CCCGGCTCCG	AGTACGGCGC	CCCGAATGGC	GTCACCGGCT	1200
GGIAACCACG	CLIGCGCGCC	TGGGCGGCGG	CCTGCCGGAT		ATGCCGACAA	1260
AGCCTGCGTG	ATCGGTCATC	ACCAACGGTG		GTTGTGCACC		
CCACCCCGGT		GTCCAGCCGA				1320
CCGGCATCAC	GTTGCCGATC				TGACCAAACC	1380
				ATGAAAATTT	AAGGGCACCA	1440
	CCCTTATCCC	ACTTGCCGTC	GGTTGCGGGT	CAGGCCCGTG	ACCAGCTCCC	1500
GCGACAAGAA	CCGTATGCCG	TCGATCTCGC	CTCGTGCCG			1539

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 851 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CTGCAGGGTG	GCGTGGATGA	CCCTCLCCC				
-		GCGTCACCGC	GGGGCAGGCC	GAGCTGACCG	CCGCCCAGGT	60
CCGGGTTGCT	GCGGCGGCCT	ACGAGACGGC	GTATGGGCTG	ACGGTGCCCC	CGCCGGTGAT	120
CGCCGAGAAC	CGTGCTGAAC	TGATGATTCT	CATACCCACC	330000000	COCCOGIGAI	120
CCCGGCGATC	GCGCTCTACC	30000000000	CAIAGCGACC	MACCICITIGG	GGCAAAACAC	180
Camommooo	GCGGTCAACG	AGGCCGAATA	CGGCGAGATG	TGGGCCCAAG	ACGCCGCCGC	240
GAIGITIGGC	TAUGUCGCGG	CGACGGCGAC	GGCGACGGCG	ACGIMECTIC	COMPCONCON	
GGCGCCGGAG	ATGACCAGCG	CGGGTGGGCT	CCTCCACCAC	2002202	CGIICGAGGA	300
CTCCGACACA	GCCCCCCC	1000100001	CCTCGAGCAG	GCCGCCGCGG	TCGAGGAGGC	360
222221222	GCCGCGA	ACCAGTTGAT	GAACAATGTG	CCCCAGGCGC	TGAAACAGTT	420
GOCCCAGCCC	ACGCAGGGCA	CCACGCCTTC	TTCCAAGCTG	GGTGGCCTTGT	CCAACACCC	
CTCGCCGCAT	CGGTCGCCGA	TCAGCAACAT	CCTCTCCC	0010000101	GGMAGACGGI	480
GACCAACTCC	COMOMOMO	TORGONACAI	GGIGICGATG	GCCAACAACC	ACATGTCGAT	540
CACCARCICG	GGIGIGICGA	TGACCAACAC	CTTGAGCTCG	ATGTTGAAGG	GCTTTGCTCC	600
GGCGGCGGCC	GCCCAGGCCG	TGCAAACCGC	GGCGCAAAAC			
GCTGGGCAGC	TCGCTGGGTT	CTTCGGGTCT			CGATGAGCTC	660
GGCGGGGGGG					ACTTGGGTCG	720
			TGGCGGAAAA	TATGCANAGT	CTGGTCGGCG	780
GAACGGTGGT	CCGGCGTAAG	GTTTACCCCC	Chiminatical	TCCCCTC	010010000	780
GAAACAGTTA	C		GITTICIGGA	TGCGGTGAAC	TTCGTCAACG	840
	•					851
						771

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 254 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CGGGCACCTG	GCGGAAATTT GGACCA ICCGTCCAGG GGCCCG ICGTAGGTCC TCGATA GTTCGAGGGC CACTCC	AGTA ATGGCTCGCA	GAGAGGAAC	CTTACTGCTG	60 120 180 240 254
(2) INFORMAT	TION FOR SEQ ID N	0:35:			
/÷\					

- (A) LENGTH: 1227 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

	GAAGCGGCCG		GAAGTCGCTG	TTGGACCAGG	AGGGACGGGA	60
CGATCTGGCG		CGGTTCAGCC	GGGGGGGTGC			
TTTCTTCGAC	GACCGGACGC	TGGATGGTGA	CCAAACCGCG		0011111111111111	120
GATCGTGGAC	CGGATGAGCG	CGCCGTATGT				180
TATTGAGAAG	CAAGGTTCAC				TCGTCGACAC	240
GATTCGTTCA	ACTGATAAAA			CCGGCTCCTG	CGCGTGCGGG	300
CCAAGACCTG				GCGCAACACG	TACGAGCACA	360
GGCCGCCGGC	1100000000000		GAGCGATGCC	TTGCACCTGA	CCGCGTGGCG	420
AGTACACGAT	2001001010	11001001100	TGAACAGCAC	CTGGGCCTGA	TATTGCGACC	480
	TTTGTCGATC		CGACCTGGGA	GAACTGCTTG	CGGAACGCGT	540
CGCTGCTCAG	CTTGGCCAAG		AGCGCTTGTC	GCGCACGCCG	TCGTGGATAC	600
CGCACAGCGC	ATTGCGAACG	ATGGTGTCCA	CATCGCGGTT	CTCCAGCGCG	TTGAGGTATC	
CCTGAATCGC	GGTTTTGGCC	GGTCCCTCCG	AGAATGTGCC	TGCCGTGTTG		660
TGCGGACCCC	GTATATGATC	GCCGCCGTCA		CAGCGCGAGG	GCTCCGTTGG	720
TGCCGATCAG	CAGCCGCTTG	TGCCGTCGCT			GCTACCACAA	780
GATATGCGGC	GGGCGGCAGC	GCCGCGTCGT		CACCTGCGGC	GGCACGCCGG	840
CGGCGCCGAG	GTCGTGGGGG	TAGTCCAGGG		CGGGGCGAAG	GCCGGTTCGG	900
ACGGCGCCGG	TCCGTTGGTG			GTGGGATGAG	GGCTCGGGGT	960
TTCTCCTAGG	_	CCGACACCGG	GGTTCGGCGA	GTGGGGACCG	GGCATTGTGG	1020
GGCAGCATCG	GTGGTGGACG	GGACCAGCTG	CTAGGGCGAC	AACCGCCCGT	CGCGTCAGCC	1080
		GAGCTCCCTA	GGCAGGCTAG	CGCAACAGCT	GCCGTCAGCT	1140
CTCAACGCGA	CGGGGCGGGC	CGCGGCGCCG	ATAATGTTGA		AACCTTAGGA	
ACGAAGGACG	GAGATTTTGT	GACGATC			ANCOT INCOM	1200
						1227

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 181 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

 THACGG LGGT	GCCGGCGCZ	VCCC-income	CEMCCCCCCC	GGGGCCGGCG GGCGGGTCCG GGCAACGGCG	120 180
					181

(2) INFORMATION FOR SEQ ID NO:37:

(A) LENGTH: 290 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:	
GCGGTGTCGG CGGATCCGGC CCGTCCTTGA ACCOUNTS	
GCGGTGTCGG CGGATCCGGC GGGTGGTTGA ACGGCAACGG CGGTGTCGGC GGCCGGGGCG GCGACGGCGT CTTTGCCGGT GCCGGCGGCC AGGCCGGCCT CGGTGGCAACG GGCGGCAATG	60
	120
	180
GCACTCAGAG CGCGACCGGC CTCGGNGGTG ACGGCGGTGA CGGCGGTGAC	240
	290
(2) INFORMATION FOR SEQ ID NO:38:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 34 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:	
GATCCAGTGG CATGGNGGGT GTCAGTGGAA GCAT	34
(2) INFORMATION FOR SEQ ID NO:39:	34
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 155 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
Timear Timear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:	
GATCGCTGCT CGTCCCCCC TTGCCGCCGA CGCCACCGGT CCCACCGTTA CCGAACAAGC	
	60
TATCCCCACC ATTGCCGCCG GNCCCACCGG CACCG	120
	155
(2) INFORMATION FOR SEQ ID NO:40:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 53 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:	
ATGGCGTTCA CGGGGCCCC GGGACCGGGC AGCCCGGNGG GGCCGGGGG TGG	53
(2) INFORMATION FOR SEQ ID NO:41:	

(C) STRANDEDNESS: single

(A) LENGTH: 132 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:	
GATCCACCGC GGGTGCAGAC GGTGCCCGCG GCGCCACCCC GACCAGCGGC GGCAACGGCG GCACCGGCGG CAACGCCGCG AACGCCACCG TCGTCGGNGG GGCCGGCGGG GCCGGCGGCA AGGGCGGCAA CG	60 120 132
(2) INFORMATION FOR SEQ ID NO:42:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 132 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:	
GATCGGCGGC CGGNACGGNC GGGGACGGCG GCAAGGGCGG NAACGGGGGC GCCGNAGCCA CCNGCCAAGA ATCCTCCGNG TCCNCCAATG GCGCGAATGG CGGACAGGGC GGCAACGGCG GCANCGGCGG CA	60 120 132
(2) INFORMATION FOR SEQ ID NO:43:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 702 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:	
CGGCACGAGG ATCGGTACCC CGCGGCATCG GCAGCTGCCG ATTCGCCGGG TTTCCCCACC	60
CGAGGAAAGC CGCTACCAGA TGGCGCTGCC GAAGTAGGGC GATCCGTTCG CGATGCCGGC ATGAACGGGC GGCATCAAAT TAGTGCAGGA ACCTTTCAGT TTAGCGACGA TAATGGCTAT	120
AGCACTAAGG AGGATGATCC GATATGACGC AGTCGCAGAC CGTGACGGTG GATCAGCAAG	180 240
AGATTITGAA CAGGGCCAAC GAGGTGGAGG CCCCGATGGC GGACCCACCG ACTGATGTCC	300
CCATCACACC GTGCGAACTC ACGGNGGNTA AAAACGCCGC CCAACAGNTG GTNTTGTCCG	360
CGACAACAT GCGGGAATAC CTGGCGGCCG GTGCCAAAGA GCGGCAGCGT CTGGCGACCT	420
GCTGCGCAA CGCGGCCAAG GNGTATGGCG AGGTTGATGA GGAGGCTGCG ACCGCGCTCG	480
ACAACGACGG CGAAGGAACT GTGCAGGCAG AATCGGCCGG GGCCGTCGGA GGGGACAGTT	540
EGGCCGAACT AACCGATACG CCGAGGGTGG CCACGGCCGG TGAACCCAAC TTCATGGATC	600
TCAAAGAAGC GGCAAGGAAG CTCGAAACGG GCGACCAAGG CGCATCGCTC GCGCACTGNG	660
GGGATGGGTG GAACACTINC ACCCTGACGC TGCAAGGCGA CG	702
(2) INFORMATION FOR SEQ ID NO:44:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 298 base pairs	
(B) TYPE: nucleic acid	

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GAAGCCGCAG	CGCTGTCGGG	CCACCTCCCC	CTCTTTTTTTT	01 magamas -	TGGCGGTGGA	
22222222		COACGIGGCG	GICAAAGCGG	CATCGCTCGG	TGGCGGTGGA	60
GGCGGCGGG	TGCCGTCGGC	GCCGTTGGGA	TCCGCGATCG	GGGGCGCCGA	ATCGGTGCGG	120
CCCGCTGGCG	CTGGTGACAT	TGCCGGCTTA	GGCCAGGGAA	cccccccc	CGGCGCGCG	
CTGGGGGGGGG	CTCCCARTCCC	11000001111	COCCAGGGAA	GGGCGG	CGGCGCGCG	180
CIGGGCGGCG	GIGGCAIGG	AATGCCGATG	GGTGCCGCGC	ATCAGGGACA	AGGGGGCGCC	240
AAGTCCAAGG	GTTCTCAGCA	GGAAGACGAG	GCGCTCTACA	CCGAGGATCC	TCGTGCCG	298

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1058 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

CGGCACGAGG	ATCGAATCGC	GTCGCCGGGA	GCACAGCGTC	GCACTGCACC	AGTGGAGGAG	60
CCATGACCTA	CTCGCCGGGT	AACCCCGGAT	ACCCGCAAGC	GCAGCCCGCA	GGCTCCTACG	120
GAGGCGTCAC	ACCCTCGTTC	GCCCACGCCG	ATGAGGGTGC	GAGCAAGCTA	CCGATGTACC	180
TGAACATCGC	GGTGGCAGTG	CTCGGTCTGG	CTGCGTACTT	CGCCAGCTTC	GGCCCAATGT	240
TCACCCTCAG				GTCCGGTGAC		300
CGGTCGGGGT	GGCTCTGCTG	GCTGCGCTGC	TTGCCGGGGT	GGTTCTGGTG	CCTAAGGCCA	360
AGAGCCATGT				CGTATTTCTG		
CGACGTTTAA	CAAGCCCAGC	GCCTATTCGA	CCGGTTGGGC	ATTGTGGGTT	GTGTTGGCTT	420
TCATCGTGTT				GGTGGAGACC		480
CCGCGCCGGC	GCCGCGGCCC	AAGTTCGACC	CGTATGGACA	GTACGGGCGG	TACGGGCAGT	540
ACGGGCAGTA	CGGGGTGCAG	CCGGGTGGGT	ACTACGGTCA	GCAGGGTGCT	CACCACCACA	600
CGGGACTGCA	GTCGCCCGGC	CCGCAGCAGT	CTCCCCACCC	TCCCGGATAT	CAGCAGGCCG	660
				ATACACTGCT		720
				GGGCCCATCC		780
		CCACCACCAC			ACGCCACCTA	840
GTTCGGCTCC					TCGCAGGCTG	900
					TCCCCCGGG	960
GGGTGTCAGC	AACCCCCCAC	GTTCCCGCGT	CCGGTCGCGC	GTGTGCGCGA	AGAGTGAACA	1020
300101CAGC	MUCUCUUAL	GMICCICGIG	CCGAATTC			1058

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 327 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CGGCACGAGA	GACCGATGCC	GCTACCCTCG	CGCAGGAGGC	AGGTAATTTC	GAGCGGATCT	60
CCGGCGACCT	GAAAACCCAG	ATCGACCAGG	TGGAGTCGAC	GGCAGGTTCG	TTGCAGGGCC	120
AGTGGCGCGG	CGCGGCGGG	ACGGCCGCCC	AGGCCGCGGT	GGTGCGCTTC	CAAGAAGCAG	180
CCAATAAGCA	GAAGCAGGAA	CTCGACGAGA	TCTCGACGAA	TATTCGTCAG	GCCGGCGTCC	240
AATACTCGAG	GGCCGACGAG	GAGCAGCAGC	AGGCGCTGTC	CTCGCAAATG	GGCTTCTGAC	300
CCGCTAATAC	GAAAAGAAAC	GGAGCAA				327

(2) INFORMATION FOR SEQ ID NO:47:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 170 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:	
CGGTCGCGAT GATGGCGTTG TCGAACGTGA CCGATTCTGT ACCGCCGTCG TTGAGATCAA CCAACAACGT GTTGGCGTCG GCAAATGTGC CGNACCCGTG GATCTCGGTG ATCTTGTTCT TCTTCATCAG GAAGTGCACA CCGGCCACCC TGCCCTCGGN TACCTTTCGG	60 120 170
(2) INFORMATION FOR SEQ ID NO:48:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 127 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:	
GATCCGGCGG CACGGGGGGT GCCGGCGGCA GCACCGCTGG CGCTGGCGGC AACGGCGGGG CCGGGGGTGG CGGCGGAACC GGTGGGTTGC TCTTCGGCAA CGGCGGTGCC GGCGGGCACG GGGCCGT	60 120 127
(2) INFORMATION FOR SEQ ID NO:49:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 81 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:	
CGGCGGCAAG GGCGGCACCG CCGGCAACGG GAGCGGCGCG GCCGGCGGCA ACGGCGGCAA	60 81
(2) INFORMATION FOR SEQ ID NO:50:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 149 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:	
GATCAGGGCT GGCCGGCTCC GGCCAGAAGG GCGGTAACGG AGGAGCTGCC GGATTGTTTG GCAACGGCGG GGCCGGNGGT GCCGGCGCGT CCAACCAAGC CGGTAACGGC GGNGCCGGCG GAAACGGTGG TGCCGGTGGG CTGATCTGG	60 120 149

(2) INFORMATION FOR SEQ ID NO:51:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 355 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

TCGAAGTACA CTATGAAAGT	GTCAATTCGA CGGCTTCCGC TAAGCGACTT	GGCCACCTGG CTGGAGGATT	TCGACGGAGC CCTGAACCTT	CATGCGCGCG GGTCGCGCAC CAAGCGCGGC	GACGGTGTCG CTGGACTGGT TTCCAGGTGA CGATAACTGA GGTTCAGCCG GGCGG	60 120 180 240 300
				carriregie	GGCGG	355

(2) INFORMATION FOR SEQ ID NO:52:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 999 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

(2) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 332 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Met His His His His His Met His Gln Val Asp Pro Asn Leu Thr Arg Arg Lys Gly Arg Leu Ala Ala Leu Ala Ile Ala Ala Met Ala Ser Ala Ser Leu Val Thr Val Ala Val Pro Ala Thr Ala Asn Ala Asp Pro 40 Glu Pro Ala Pro Pro Val Pro Thr Thr Ala Ala Ser Pro Pro Ser Thr Ala Ala Ala Pro Pro Ala Pro Ala Thr Pro Val Ala Pro Pro Pro 75 Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro Gly Asp Pro Asn Ala Ala 90 Pro Pro Pro Ala Asp Pro Asn Ala Pro Pro Pro Pro Val Ile Ala Pro 105 Asn Ala Pro Gln Pro Val Arg Ile Asp Asn Pro Val Gly Gly Phe Ser 120 Phe Ala Leu Pro Ala Gly Trp Val Glu Ser Asp Ala Ala His Phe Asp 135 Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr Gly Asp Pro Pro Phe Pro 140 150 Gly Gln Pro Pro Pro Val Ala Asn Asp Thr Arg Ile Val Leu Gly Arg 155 170 Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu Ala Thr Asp Ser Lys Ala 185 Ala Ala Arg Leu Gly Ser Asp Met Gly Glu Phe Tyr Met Pro Tyr Pro 200 Gly Thr Arg Ile Asn Gln Glu Thr Val Ser Leu Asp Ala Asn Gly Val 215 Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys Phe Ser Asp Pro Ser Lys 220 Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn 235 250 Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly 265 Thr Ala Asn Asn Pro Val Asp Lys Gly Ala Ala Lys Ala Leu Ala Glu 280 Ser Ile Arg Pro Leu Val Ala Pro Pro Pro Ala Pro Ala Pro Ala Pro 295 Ala Glu Pro Ala Pro Ala Pro Ala Gly Glu Val Ala Pro Thr 300 315 Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu Pro Ala 325

(2) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Asp Pro Val Asp Ala Val Ile Asn Thr Thr Xaa Asn Tyr Gly Gln Val

94

1 5 10 15 Val Ala Ala Leu 20

- (2) INFORMATION FOR SEQ ID NO:55:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ala Val Glu Ser Gly Met Leu Ala Leu Gly Thr Pro Ala Pro Ser
1 5 10

- (2) INFORMATION FOR SEQ ID NO:56:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Glu Ala Ala Lys 1 5 10 10 15

- (2) INFORMATION FOR SEQ ID NO:57:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Tyr Tyr Trp Cys Pro Gly Gln Pro Phe Asp Pro Ala Trp Gly Pro 1 5 10 15

- (2) INFORMATION FOR SEQ ID NO:58:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 14 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Asp Ile Gly Ser Glu Ser Thr Glu Asp Gln Gln Xaa Ala Val

- (2) INFORMATION FOR SEQ ID NO:59:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

Ala Glu Glu Ser Ile Ser Thr Xaa Glu Xaa Ile Val Pro 5 10

- (2) INFORMATION FOR SEQ ID NO:60:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

- (2) INFORMATION FOR SEQ ID NO:61:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Ala Pro Lys Thr Tyr Xaa Glu Glu Leu Lys Gly Thr Asp Thr Gly

15

- (2) INFORMATION FOR SEQ ID NO:62:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln Gln Thr Ser

Leu Leu Asn Asn Leu Ala Asp Pro Asp Val Ser Phe Ala Asp 20 25 30

(2) INFORMATION FOR SEQ ID NO:63:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 187 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Thr Gly Ser Leu Asn Gln Thr His Asn Arg Arg Ala Asn Glu Arg Lys

1 10 15

Asn Thr Thr Met Lys Met Val Lys Ser Ile Ala Ala Gly Leu Thr Ala

Ala Ala Ala Ile Gly Ala Ala Ala Gly Val Thr Ser Ile Met Ala
35 40 45

Gly Gly Pro Val Val Tyr Gln Met Gln Pro Val Val Phe Gly Ala Pro

Leu Pro Leu Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln 65 70 75 80

Leu Thr Ser Leu Leu Asn Ser Leu Ala Asp Pro Asn Val Ser Phe Ala 85 90 95

Asn Lys Gly Ser Leu Val Glu Gly Gly Ile Gly Gly Thr Glu Ala Arg

Ile Ala Asp His Lys Leu Lys Lys Ala Ala Glu His Gly Asp Leu Pro

Leu Ser Phe Ser Val Thr Asn Ile Gln Pro Ala Ala Ala Gly Ser Ala
130 135 140

Gln Asn Val Thr Phe Val Asn Gln Gly Gly Trp Met Leu Ser Arg Ala
165 170 175

Ser Ala Met Glu Leu Leu Gln Ala Ala Gly Xaa

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 148 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

Asp Glu Val Thr Val Glu Thr Thr Ser Val Phe Arg Ala Asp Phe Leu

1 5 10 15

Ser Glu Leu Asp Ala Pro Ala Gln Ala Gly Thr Glu Ser Ala Val Ser 20 25 30

Gly Val Glu Gly Leu Pro Pro Gly Ser Ala Leu Leu Val Val Lys Arg
35 40 45

Gly Pro Asn Ala Gly Ser Arg Phe Leu Leu Asp Gln Ala Ile Thr Ser

 Ala
 Gly
 Arg
 His
 Pro
 Asp
 Ser
 Asp
 Ile
 Phe
 Leu
 Asp
 Asp
 Val
 Thr
 Val

 Ser
 Arg
 His
 Ala
 Glu
 Phe
 Arg
 Leu
 Ala
 Ala
 Ala
 Arg
 Leu
 Ala
 Ala
 Arg
 Leu
 Arg
 A

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 230 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Thr Ser Asn Arg Pro Ala Arg Arg Gly Arg Arg Ala Pro Arg Asp Thr Gly Pro Asp Arg Ser Ala Ser Leu Ser Leu Val Arg His Arg Arg Gln 25 Gln Arg Asp Ala Leu Cys Leu Ser Ser Thr Gln Ile Ser Arg Gln Ser 40 Asn Leu Pro Pro Ala Ala Gly Gly Ala Ala Asn Tyr Ser Arg Arg Asn 55 Phe Asp Val Arg Ile Lys Ile Phe Met Leu Val Thr Ala Val Val Leu 70 75 Leu Cys Cys Ser Gly Val Ala Thr Ala Ala Pro Lys Thr Tyr Cys Glu 85 Glu Leu Lys Gly Thr Asp Thr Gly Gln Ala Cys Gln Ile Gln Met Ser 105 Asp Pro Ala Tyr Asn Ile Asn Ile Ser Leu Pro Ser Tyr Tyr Pro Asp 120 125 Gln Lys Ser Leu Glu Asn Tyr Ile Ala Gln Thr Arg Asp Lys Phe Leu 135 140 Ser Ala Ala Thr Ser Ser Thr Pro Arg Glu Ala Pro Tyr Glu Leu Asn 150 155 Ile Thr Ser Ala Thr Tyr Gln Ser Ala Ile Pro Pro Arg Gly Thr Gln 165 170 Ala Val Val Leu Xaa Val Tyr His Asn Ala Gly Gly Thr His Pro Thr 185 Thr Thr Tyr Lys Ala Phe Asp Trp Asp Gln Ala Tyr Arg Lys Pro Ile 200 Thr Tyr Asp Thr Leu Trp Gln Ala Asp Thr Asp Pro Leu Pro Val Val 215 Phe Pro Ile Val Ala Arg

(2) INFORMATION FOR SEQ ID NO:66:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 132 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

Thr Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gln Gly Phe

1 5 10 15

Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Na Glo Glo Gly Fle

Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser

Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly
35 40

Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val 50 55 60

Val Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val

Ile Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala 85 90 95

Asp Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp

Gln Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu

Gly Pro Pro Ala 130

(2) INFORMATION FOR SEQ ID NO:67:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 100 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

Val Pro Leu Arg Ser Pro Ser Met Ser Pro Ser Lys Cys Leu Ala Ala 1 5 10

Ala Gln Arg Asn Pro Val Ile Arg Arg Arg Leu Ser Asn Pro Pro
20 25 30

Pro Arg Lys Tyr Arg Ser Met Pro Ser Pro Ala Thr Ala Ser Ala Gly
35 40 45

Met Ala Arg Val Arg Arg Arg Ala Ile Trp Arg Gly Pro Ala Thr Xaa 50 55 60

Ser Ala Gly Met Ala Arg Val Arg Arg Trp Xaa Val Met Pro Xaa Val 65 70 75 80

Ile Gln Ser Thr Xaa Ile Arg Xaa Xaa Gly Pro Phe Asp Asn Arg Gly
85 90 95

Ser Glu Arg Lys

(2) INFORMATION FOR SEQ ID NO:68:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 163 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

Met Thr Asp Asp Ile Leu Leu Ile Asp Thr Asp Glu Arg Val Arg Thr Leu Thr Leu Asn Arg Pro Gln Ser Arg Asn Ala Leu Ser Ala Ala Leu 25 Arg Asp Arg Phe Phe Ala Xaa Leu Xaa Asp Ala Glu Xaa Asp Asp Asp Ile Asp Val Val Ile Leu Thr Gly Ala Asp Pro Val Phe Cys Ala Gly Leu Asp Leu Lys Val Ala Gly Arg Ala Asp Arg Ala Ala Gly His Leu 75 Thr Ala Val Gly Gly His Asp Gln Ala Gly Asp Arg Arg Asp Gln Arg 90 Arg Arg Gly His Arg Arg Ala Arg Thr Gly Ala Val Leu Arg His Pro 105 Asp Arg Leu Arg Ala Arg Pro Leu Arg Arg His Pro Arg Pro Gly Gly 120 Ala Ala His Leu Gly Thr Gln Cys Val Leu Ala Ala Lys Gly Arg 135 His Arg Xaa Gly Pro Val Asp Glu Pro Asp Arg Arg Leu Pro Val Arg 145 150 155 Asp Arg Arg

(2) INFORMATION FOR SEQ ID NO:69:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 344 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

 Met
 Lys
 Phe
 Val
 Ass
 His
 Ile
 Glu
 Pro
 Val
 Ala
 Pro
 Arg
 Arg
 Ala
 Gly
 Arg
 Ala
 Gly
 Arg
 Ala
 Gly
 Arg
 Ala
 Gly
 Arg
 Ala
 Arg
 Arg
 Arg
 Arg
 Alu
 Arg
 Arg</th

WO 99/42076 PCT/US99/03268

100

Gly Asp Pro Asn Ala Pro Tyr Val Ala Trp Ala Ala Gly Thr Gly Thr 120 Pro Ala Gly Pro Pro Ala Pro Phe Gly Pro Asp Val Ala Ala Glu Tyr 135 Leu Gly Thr Ala Val Gln Phe His Phe Ile Ala Arg Leu Val Leu Val 140 150 155 Leu Leu Asp Glu Thr Phe Leu Pro Gly Gly Pro Arg Ala Gln Gln Leu 165 170 Met Arg Arg Ala Gly Gly Leu Val Phe Ala Arg Lys Val Arg Ala Glu 185 His Arg Pro Gly Arg Ser Thr Arg Arg Leu Glu Pro Arg Thr Leu Pro 200 Asp Asp Leu Ala Trp Ala Thr Pro Ser Glu Pro Ile Ala Thr Ala Phe 215 Ala Ala Leu Ser His His Leu Asp Thr Ala Pro His Leu Pro Pro 230 235 Thr Arg Gln Val Val Arg Arg Val Val Gly Ser Trp His Gly Glu Pro 245 250 Met Pro Met Ser Ser Arg Trp Thr Asn Glu His Thr Ala Glu Leu Pro Ala Asp Leu His Ala Pro Thr Arg Leu Ala Leu Leu Thr Gly Leu Ala Pro His Gln Val Thr Asp Asp Val Ala Ala Ala Arg Ser Leu Leu 295 Asp Thr Asp Ala Ala Leu Val Gly Ala Leu Ala Trp Ala Ala Phe Thr 315 Ala Ala Arg Arg Ile Gly Thr Trp Ile Gly Ala Ala Ala Glu Gly Gln 330 Val Ser Arg Gln Asn Pro Thr Gly 340

(2) INFORMATION FOR SEQ ID NO:70:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 485 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

Asp Gln Tyr Glu Pro Gly Ser Ser Arg Arg Trp Ala Glu Arg Phe Ala 120 Thr Leu Leu Arg Asn Leu Glu Phe Leu Pro Asn Ser Pro Thr Leu Met 135 Asn Ser Gly Thr Asp Leu Gly Leu Leu Ala Gly Cys Phe Val Leu Pro 150 155 Ile Glu Asp Ser Leu Gln Ser Ile Phe Ala Thr Leu Gly Gln Ala Ala 170 Glu Leu Gln Arg Ala Gly Gly Gly Thr Gly Tyr Ala Phe Ser His Leu 185 Arg Pro Ala Gly Asp Arg Val Ala Ser Thr Gly Gly Thr Ala Ser Gly 200 Pro Val Ser Phe Leu Arg Leu Tyr Asp Ser Ala Ala Gly Val Val Ser 215 Met Gly Gly Arg Arg Gly Ala Cys Met Ala Val Leu Asp Val Ser 230 His Pro Asp Ile Cys Asp Phe Val Thr Ala Lys Ala Glu Ser Pro Ser 235 250 Glu Leu Pro His Phe Asn Leu Ser Val Gly Val Thr Asp Ala Phe Leu 265 Arg Ala Val Glu Arg Asn Gly Leu His Arg Leu Val Asn Pro Arg Thr 280 Gly Lys Ile Val Ala Arg Met Pro Ala Ala Glu Leu Phe Asp Ala Ile 295 Cys Lys Ala Ala His Ala Gly Gly Asp Pro Gly Leu Val Phe Leu Asp 300 310 315 Thr Ile Asn Arg Ala Asn Pro Val Pro Gly Arg Gly Arg Ile Glu Ala 325 330 Thr Asn Pro Cys Gly Glu Val Pro Leu Leu Pro Tyr Glu Ser Cys Asn 345 Leu Gly Ser Ile Asn Leu Ala Arg Met Leu Ala Asp Gly Arg Val Asp 360 Trp Asp Arg Leu Glu Glu Val Ala Gly Val Ala Val Arg Phe Leu Asp 375 Asp Val Ile Asp Val Ser Arg Tyr Pro Phe Pro Glu Leu Gly Glu Ala 390 Ala Arg Ala Thr Arg Lys Ile Gly Leu Gly Val Met Gly Leu Ala Glu 395 405 410 Leu Leu Ala Ala Leu Gly Ile Pro Tyr Asp Ser Glu Glu Ala Val Arg 425 Leu Ala Thr Arg Leu Met Arg Arg Ile Gln Gln Ala Ala His Thr Ala 440 Ser Arg Arg Leu Ala Glu Glu Arg Gly Ala Phe Pro Ala Phe Thr Asp 455 Ser Arg Phe Ala Arg Ser Gly Pro Arg Arg Asn Ala Gln Val Thr Ser 460 475 Val Ala Pro Thr Gly 485

(2) INFORMATION FOR SEQ ID NO:71:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 267 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single

102

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

Gly Val Ile Val Leu Asp Leu Glu Pro Arg Gly Pro Leu Pro Thr Glu Ile Tyr Trp Arg Arg Gly Leu Ala Leu Gly Ile Ala Val Val 25 Val Gly Ile Ala Val Ala Ile Val Ile Ala Phe Val Asp Ser Ser Ala Gly Ala Lys Pro Val Ser Ala Asp Lys Pro Ala Ser Ala Gln Ser His Pro Gly Ser Pro Ala Pro Gln Ala Pro Gln Pro Ala Gly Gln Thr Glu 70 Gly Asn Ala Ala Ala Pro Pro Gln Gly Gln Asn Pro Glu Thr Pro 90 Thr Pro Thr Ala Ala Val Gln Pro Pro Pro Val Leu Lys Glu Gly Asp 100 105 Asp Cys Pro Asp Ser Thr Leu Ala Val Lys Gly Leu Thr Asn Ala Pro 120 Gln Tyr Tyr Val Gly Asp Gln Pro Lys Phe Thr Met Val Val Thr Asn 135 Ile Gly Leu Val Ser Cys Lys Arg Asp Val Gly Ala Ala Val Leu Ala Ala Tyr Val Tyr Ser Leu Asp Asn Lys Arg Leu Trp Ser Asn Leu Asp 170 Cys Ala Pro Ser Asn Glu Thr Leu Val Lys Thr Phe Ser Pro Gly Glu 180 185 Gln Val Thr Thr Ala Val Thr Trp Thr Gly Met Gly Ser Ala Pro Arg 200 Cys Pro Leu Pro Arg Pro Ala Ile Gly Pro Gly Thr Tyr Asn Leu Val 215 Val Gln Leu Gly Asn Leu Arg Ser Leu Pro Val Pro Phe Ile Leu Asn 230 Gln Pro Pro Pro Pro Pro Gly Pro Val Pro Ala Pro Gly Pro Ala Gln Ala Pro Pro Pro Glu Ser Pro Ala Gln Gly Gly

(2) INFORMATION FOR SEQ ID NO:72:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 97 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

Leu Ile Ser Thr Gly Lys Ala Ser His Ala Ser Leu Gly Val Gln Val 1 1×10^{-5} From Asn Asp Lys Asp Thr Pro Gly Ala Lys Ile Val Glu Val Val Ala 20 1×10^{-5} From Asp Ala Ala Ala Asn Ala Gly Val Pro Lys Gly Val Val Val Thr 35 1×10^{-5} From Asp Ala Ala Ala Asp Ala Gly Val Pro Lys Gly Val Val Val Thr

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 364 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Gly Ala Ala Val Ser Leu Leu Ala Ala Gly Thr Leu Val Leu Thr Ala 10 Cys Gly Gly Gly Thr Asn Ser Ser Ser Ser Gly Ala Gly Gly Thr Ser Gly Ser Val His Cys Gly Gly Lys Lys Glu Leu His Ser Ser Gly Ser 40 Thr Ala Gln Glu Asn Ala Met Glu Gln Phe Val Tyr Ala Tyr Val Arg 55 Ser Cys Pro Gly Tyr Thr Leu Asp Tyr Asn Ala Asn Gly Ser Gly Ala Gly Val Thr Gln Phe Leu Asn Asn Glu Thr Asp Phe Ala Gly Ser Asp 85 90 Val Pro Leu Asn Pro Ser Thr Gly Gln Pro Asp Arg Ser Ala Glu Arg 105 Cys Gly Ser Pro Ala Trp Asp Leu Pro Thr Val Phe Gly Pro Ile Ala 120 125 Ile Thr Tyr Asn Ile Lys Gly Val Ser Thr Leu Asn Leu Asp Gly Pro 135 140 Thr Thr Ala Lys Ile Phe Asn Gly Thr Ile Thr Val Trp Asn Asp Pro 150 155 Gln Ile Gln Ala Leu Asn Ser Gly Thr Asp Leu Pro Pro Thr Pro Ile 165 170 Ser Val Ile Phe Arg Ser Asp Lys Ser Gly Thr Ser Asp Asn Phe Gln 185 Lys Tyr Leu Asp Gly Val Ser Asn Gly Ala Trp Gly Lys Gly Ala Ser 195 200 Glu Thr Phe Ser Gly Gly Val Gly Val Gly Ala Ser Gly Asn Asn Gly 215 220 Thr Ser Ala Leu Leu Gln Thr Thr Asp Gly Ser Ile Thr Tyr Asn Glu 230 235 Trp Ser Phe Ala Val Gly Lys. Gln Leu Asn Met Ala Gln Ile Ile Thr 245 250 Ser Ala Gly Pro Asp Pro Val Ala Ile Thr Thr Glu Ser Val Gly Lys Thr Ile Ala Gly Ala Lys Ile Met Gly Gln Gly Asn Asp Leu Val Leu 280 Asp Thr Ser Ser Phe Tyr Arg Pro Thr Gln Pro Gly Ser Tyr Pro Ile

 Val
 Leu
 Ala
 Thr
 Tyr
 Glu
 Ile
 Val
 Cys
 Ser
 Lys
 Tyr
 Pro
 Asp
 Ala
 Thr

 305
 -

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 309 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

Gln Ala Ala Gly Arg Ala Val Arg Arg Thr Gly His Ala Glu Asp 5 10 Gln Thr His Gln Asp Arg Leu His His Gly Cys Arg Arg Ala Ala Val 25 Val Val Arg Gln Asp Arg Ala Ser Val Ser Ala Thr Ser Ala Arg Pro 40 Pro Arg Arg His Pro Ala Gln Gly His Arg Arg Arg Val Ala Pro Ser Gly Gly Arg Arg Pro His Pro His His Val Gln Pro Asp Asp Arg 70 75 Arg Asp Arg Pro Ala Leu Leu Asp Arg Thr Gln Pro Ala Glu His Pro 90 Asp Pro His Arg Arg Gly Pro Ala Asp Pro Gly Arg Val Arg Gly Arg 105 Gly Arg Leu Arg Arg Val Asp Asp Gly Arg Leu Gln Pro Asp Arg Asp 120 Ala Asp His Gly Ala Pro Val Arg Gly Arg Gly Pro His Arg Gly Val Gln His Arg Gly Gly Pro Val Phe Val Arg Arg Val Pro Gly Val Arg 150 155 Cys Ala His Arg Arg Gly His Arg Arg Val Ala Ala Pro Gly Gln Gly 170 Asp Val Leu Arg Ala Gly Leu Arg Val Glu Arg Leu Arg Pro Val Ala 185 Ala Val Glu Asn Leu His Arg Gly Ser Gln Arg Ala Asp Gly Arg Val 200 Phe Arg Pro Ile Arg Arg Gly Ala Arg Leu Pro Ala Arg Arg Ser Arg 215 Ala Gly Pro Gln Gly Arg Leu His Leu Asp Gly Ala Gly Pro Ser Pro 230 Leu Pro Ala Arg Ala Gly Gln Gln Pro Ser Ser Ala Gly Gly Arg 250 Arg Ala Gly Gly Ala Glu Arg Ala Asp Pro Gly Gln Arg Gly Arg His His Gln Gly Gly His Asp Pro Gly Arg Gln Gly Ala Gln Arg Gly Thr

105

275 280 285

Ala Gly Val Ala His Ala Ala Gly Pro Arg Arg Ala Ala Val Arg
290 295 300

Asn Arg Pro Arg Arg
305

(2) INFORMATION FOR SEQ ID NO:75:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 580 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

Ser Ala Val Trp Cys Leu Asn Gly Phe Thr Gly Arg His Arg His Gly Arg Cys Arg Val Arg Ala Ser Gly Trp Arg Ser Ser Asn Arg Trp Cys 25 Ser Thr Thr Ala Asp Cys Cys Ala Ser Lys Thr Pro Thr Gln Ala Ala Ser Pro Leu Glu Arg Arg Phe Thr Cys Cys Ser Pro Ala Val Gly Cys 55 Arg Phe Arg Ser Phe Pro Val Arg Arg Leu Ala Leu Gly Ala Arg Thr Ser Arg Thr Leu Gly Val Arg Arg Thr Leu Ser Gln Trp Asn Leu Ser 85 90 Pro Arg Ala Gln Pro Ser Cys Ala Val Thr Val Glu Ser His Thr His 105 Ala Ser Pro Arg Met Ala Lys Leu Ala Arg Val Val Gly Leu Val Gln 120 Glu Glu Gln Pro Ser Asp Met Thr Asn His Pro Arg Tyr Ser Pro Pro 135 140 Pro Gln Gln Pro Gly Thr Pro Gly Tyr Ala Gln Gly Gln Gln Gln Thr 150 155 Tyr Ser Gln Gln Phe Asp Trp Arg Tyr Pro Pro Ser Pro Pro Pro Gln 165 170 Pro Thr Gln Tyr Arg Gln Pro Tyr Glu Ala Leu Gly Gly Thr Arg Pro 180 185 Gly Leu Ile Pro Gly Val Ile Pro Thr Met Thr Pro Pro Pro Gly Met 200 Val Arg Gln Arg Pro Arg Ala Gly Met Leu Ala Ile Gly Ala Val Thr 215 220 Ile Ala Val Val Ser Ala Gly Ile Gly Gly Ala Ala Ala Ser Leu Val 230 235 Gly Phe Asn Arg Ala Pro Ala Gly Pro Ser Gly Gly Pro Val Ala Ala 245 250 Ser Ala Ala Pro Ser Ile Pro Ala Ala Asn Met Pro Pro Gly Ser Val 260 Glu Gln Val Ala Ala Lys Val Val Pro Ser Val Val Met Leu Glu Thr 280 Asp Leu Gly Arg Gln Ser Glu Glu Gly Ser Gly Ile Ile Leu Ser Ala 295 300 Glu Gly Leu Ile Leu Thr Asn Asn His Val Ile Ala Ala Ala Lys

30.	5				310)				315					
Pro	o Pro	o Lei	u Gly	/ Ser	Pro) Dro	· 10		· mi-	- m)-:) 				320
				325		, ,,	PIC	г	ini	Thi	· val	Thr	Phe	Ser	Asp
Gly	Arc	Th:	r Ala			. mL.			330) 				335	
•		,	r Ala 340	, ,,	, Elle	1111	val	. val	. Gly	/ Ala	Asp	Pro	Thr	Ser	Asp
T1 e	Δ1:	17-1						345	i				350		
	- 1120	355	l Val	. Arg	val	. Gin	Gly	Val	Ser	Gly	' Leu	Thr	Pro	Ile	Ser
			•				360					300			
nec	227	, Sei	: Ser	Ser	Asp	Leu	Arg	Val	Gly	Gln	Pro	Val	Leu	Ala	Ile
						3/5					200				
GTÀ	y Ser	Pro	Leu	Gly	Leu	Glu	Gly	Thr	Val	Thr	Thr	Glv	Ile	Val	Ser
					ンプロ					205					
Ala	Leu	Asn	Arg	Pro	Val	Ser	Thr	Thr	Glv	Glu	Ala	Glv	Δen	Gl n	700
				*03					410						
Thr	Val	Leu	Asp	Ala	Ile	Gln	Thr	Asp	Ala	Δla	Tla	y c.m	D	415	
			720					475							
Ser	Gly	Gly	Ala	Leu	Val	Asn	Met	Agn	A1 =	G1n	T	77- 7	430		_
							44()								
Ser	Ala	Ile	Ala	Thr	Leu	Glv	Δla	Aen	Ca=	31-	.	445			
	450					455	ALG	Asp	ser	ALA	ASD	Ala	Gln	Ser	Gly
Ser	Ile	Gly	Leu	Glv	Dhe	Ala	T1.	D	17- 7	_	460				
465		•		1	470	Aud	116	PIO	val	Asp	Gln	Ala	Lys	Arg	Ile
Ala	Asp	Glu	Len	Tla	270	Th.	61			475					480
	•		Leu	485	261	1111	GIÀ	rys	Ala	Ser	His	Ala	Ser	Leu	Gly
Val	Gln	Val	Thr		7	T	_		490					495	
-			Thr 500	ASII	ASP	гÀг	ASD	Thr	Pro	Gly	Ala	Lys	Ile	Val	Glu
			300					505							
	val	515	Gly	GTĀ	Ala	Ala	Ala	Asn	Ala	Gly	Val	Pro	Lys	Gly	Val
							520								
val	530	Inr	Lys	Va⊥	Asp	Asp	Arg	Pro	Ile	Asn	Ser	Ala	Asp	Ala	Leu
						3.33					E 4 A				
val	ALA	Ala	Val	Arg	Ser	Lys	Ala	Pro	Gly	Ala	Thr	Val	Ala	Len	Thr
					220										
Phe	Gln	qaA	Pro	Ser	Gly	Gly	Ser	Arg	Thr	Val	Gln	Va 1	Thr.	Tan	200 21
				565		-		_	570			- 41	- 4+1	575	GTÀ
Lys	Ala	Glu	Gln											۵/۵	
			580												

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 233 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

 Met
 Asn
 Asp
 Gly
 Lys
 Arg
 Ala
 Val
 Thr
 Ser
 Ala
 Val
 Leu
 Jeu
 Leu
 Jeu
 J

70 75 Leu Gly Ile Thr Ser Ala Asp Val Asp Val Arg Ala Asn Pro Leu Ala 85 90 Ala Lys Gly Val Cys Thr Tyr Asn Asp Glu Gln Gly Val Pro Phe Arg 105 Val Gln Gly Asp Asn Ile Ser Val Lys Leu Phe Asp Asp Trp Ser Asn 120 Leu Gly Ser Ile Ser Glu Leu Ser Thr Ser Arg Val Leu Asp Pro Ala 135 Ala Gly Val Thr Gln Leu Leu Ser Gly Val Thr Asn Leu Gln Ala Gln 150 155 Gly Thr Glu Val Ile Asp Gly Ile Ser Thr Thr Lys Ile Thr Gly Thr 165 170 Ile Pro Ala Ser Ser Val Lys Met Leu Asp Pro Gly Ala Lys Ser Ala 180 185 Arg Pro Ala Thr Val Trp Ile Ala Gln Asp Gly Ser His His Leu Val 200 Arg Ala Ser Ile Asp Leu Gly Ser Gly Ser Ile Gln Leu Thr Gln Ser 215 Lys Trp Asn Glu Pro Val Asn Val Asp 230

(2) INFORMATION FOR SEQ ID NO:77:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 66 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

 Val
 1le
 Asp
 1le
 Gly
 Thr
 Ser
 Pro
 Thr
 Ser
 Trp
 Glu
 Gln
 Ala
 Ala
 Ala
 Ala
 Ala
 Ala
 Arg
 Ala
 Arg
 Ala
 Arg
 Ala
 Arg
 Ala
 Arg
 Ala
 Ala
 Ala
 Arg
 A

(2) INFORMATION FOR SEQ ID NO:78:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 69 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

Val Pro Pro Ala Pro Pro Leu Pro Pro Leu Pro Pro Ser Pro Ile Ser 1 5 10 15

 Cys
 Ala
 Ser
 Pro
 Ser
 Pro
 Pro
 Pro
 Leu
 Pro
 Pro
 Ala
 Pro
 Val
 Ala

 Pro
 Gly
 Fro
 Met
 Pro
 Pro
 Leu
 Asp
 Pro
 Trp
 Pro
 Ala
 Pro
 P

(2) INFORMATION FOR SEQ ID NO:79:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 355 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Met Ser Asn Ser Arg Arg Ser Leu Arg Trp Ser Trp Leu Leu Ser Val Leu Ala Ala Val Gly Leu Gly Leu Ala Thr Ala Pro Ala Gln Ala 25 Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe Ala Asp Phe Pro Ala Leu 40 Pro Leu Asp Pro Ser Ala Met Val Ala Gln Val Ala Pro Gln Val Val Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn Ala Val Gly Ala Gly Thr 70 Gly Ile Val Ile Asp Pro Asn Gly Val Val Leu Thr Asn Asn His Val Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly Gln 105 Thr Tyr Gly Val Asp Val Val Gly Tyr Asp Arg Thr Gln Asp Val Ala 120 Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala Ile Gly 135 Gly Gly Val Ala Val Gly Glu Pro Val Val Ala Met Gly Asn Ser Gly 150 Gly Gln Gly Gly Thr Pro Arg Ala Val Pro Gly Arg Val Val Ala Leu 170 Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Glu Glu Thr 185 Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp Ser 200 Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn Thr 215 Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gln Gly Phe Ala 230 235 Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser Gly 245 250 Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly Leu 265 Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val Val 280

Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val Ile
290
Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala Asp
305
Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp Gln
325
Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu Gly
365
Pro Pro Ala
355

(2) INFORMATION FOR SEQ ID NO:80:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 205 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

Ser Pro Lys Pro Asp Ala Glu Glu Gln Gly Val Pro Val Ser Pro Thr Ala Ser Asp Pro Ala Leu Leu Ala Glu Ile Arg Gln Ser Leu Asp Ala Thr Lys Gly Leu Thr Ser Val His Val Ala Val Arg Thr Thr Gly Lys 40 Val Asp Ser Leu Leu Gly Ile Thr Ser Ala Asp Val Asp Val Arg Ala Asn Pro Leu Ala Ala Lys Gly Val Cys Thr Tyr Asn Asp Glu Gln Gly 75 Val Pro Phe Arg Val Gln Gly Asp Asn Ile Ser Val Lys Leu Phe Asp 90 Asp Trp Ser Asn Leu Gly Ser Ile Ser Glu Leu Ser Thr Ser Arg Val Leu Asp Pro Ala Ala Gly Val Thr Gln Leu Leu Ser Gly Val Thr Asn 120 Leu Gln Ala Gln Gly Thr Glu Val Ile Asp Gly Ile Ser Thr Thr Lys 135 Ile Thr Gly Thr Ile Pro Ala Ser Ser Val Lys Met Leu Asp Pro Gly 150 155 Ala Lys Ser Ala Arg Pro Ala Thr Val Trp Ile Ala Gln Asp Gly Ser 170 His His Leu Val Arg Ala Ser Ile Asp Leu Gly Ser Gly Ser Ile Gln 185 Leu Thr Gln Ser Lys Trp Asn Glu Pro Val Asn Val Asp 200

(2) INFORMATION FOR SEQ ID NO:81:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 286 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

Gly Asp Ser Phe Trp Ala Ala Ala Asp Gln Met Ala Arg Gly Phe Val 10 Leu Gly Ala Thr Ala Gly Arg Thr Thr Leu Thr Gly Glu Gly Leu Gln 25 His Ala Asp Gly His Ser Leu Leu Leu Asp Ala Thr Asn Pro Ala Val 40 Val Ala Tyr Asp Pro Ala Phe Ala Tyr Glu Ile Gly Tyr Ile Xaa Glu 55 Ser Gly Leu Ala Arg Met Cys Gly Glu Asn Pro Glu Asn Ile Phe Phe 70 Tyr Ile Thr Val Tyr Asn Glu Pro Tyr Val Gln Pro Pro Glu Pro Glu Asn Phe Asp Pro Glu Gly Val Leu Gly Gly Ile Tyr Arg Tyr His Ala 105 Ala Thr Glu Gln Arg Thr Asn Lys Xaa Gln Ile Leu Ala Ser Gly Val 120 Ala Met Pro Ala Ala Leu Arg Ala Ala Gln Met Leu Ala Ala Glu Trp 135 Asp Val Ala Ala Asp Val Trp Ser Val Thr Ser Trp Gly Glu Leu Asn 155 Arg Asp Gly Val Val Ile Glu Thr Glu Lys Leu Arg His Pro Asp Arg 170 Pro Ala Gly Val Pro Tyr Val Thr Arg Ala Leu Glu Asn Ala Arg Gly 185 Pro Val Ile Ala Val Ser Asp Trp Met Arg Ala Val Pro Glu Gln Ile 200 Arg Pro Trp Val Pro Gly Thr Tyr Leu Thr Leu Gly Thr Asp Gly Phe 215 Gly Phe Ser Asp Thr Arg Pro Ala Gly Arg Arg Tyr Phe Asn Thr Asp 230 235 Ala Glu Ser Gln Val Gly Arg Gly Phe Gly Arg Gly Trp Pro Gly Arg 250 Arg Val Asn Ile Asp Pro Phe Gly Ala Gly Arg Gly Pro Pro Ala Gln 265 Leu Pro Gly Phe Asp Glu Gly Gly Gly Leu Arg Pro Xaa Lys

(2) INFORMATION FOR SEQ ID NO:82:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 173 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

35 40 Asn His Ala Met Met Leu Val Gln His Leu Leu Asp Arg Asp Leu Arg Val Glu Ile Pro Gly Val Asp Thr Val Arg Asn Gln Phe Asp Arg Pro Arg Glu Ala Leu Ala Leu Asp Gln Glu Arg Thr Val Thr Asp 90 Gln Val Gly Arg Leu Thr Ala Val Ala Arg Asp Glu Gly Asp Phe Leu 105 Gly Glu Gln Phe Met Gln Trp Phe Leu Gln Glu Gln Ile Glu Glu Val 120 Ala Leu Met Ala Thr Leu Val Arg Val Ala Asp Arg Ala Gly Ala Asn 135 140 Leu Phe Glu Leu Glu Asn Phe Val Ala Arg Glu Val Asp Val Ala Pro 150 155 Ala Ala Ser Gly Ala Pro His Ala Ala Gly Gly Arg Leu

(2) INFORMATION FOR SEQ ID NO:83:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 107 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

(2) INFORMATION FOR SEQ ID NO:84:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 125 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Val Leu Ser Val Pro Val Gly Asp Gly Phe Trp Xaa Arg Val Val Asn 1 5 10 10 15 Pro Leu Gly Gln Pro Ile Asp Gly Arg Gly Asp Val Asp Ser Asp Thr

 Arg
 Arg
 Ala
 Leu
 Glu
 Leu
 Gln
 Ala
 Pro
 Ser
 Val
 Val
 Xaa
 Arg
 Gln
 Gly
 Gly
 Gly
 Arg
 Gln
 Gly
 Ile
 Lys
 Ala
 Ile
 Asp
 Ala
 Mec
 Thr

 Val
 Lys
 Glu
 Pro
 Leu
 Xaa
 Thr
 Gly
 Ile
 Lys
 Ala
 Ile
 Asp
 Ala
 Mec
 Thr

 Fro
 Ile
 Gly
 Arg
 Gly
 Gln
 Arg
 Gly
 Ile
 Ile
 Gly
 Asp
 Arg
 Ile
 Ile
 Ile
 Ile
 Arg
 Ile
 I

(2) INFORMATION FOR SEQ ID NO:85:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 117 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

(2) INFORMATION FOR SEQ ID NO:86:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 103 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

(2) INFORMATION FOR SEQ ID NO:87:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 88 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

 Val
 Glu
 Try
 Leu
 Glu
 Ile
 Glu
 Try
 Arg
 Gly
 Met
 Leu
 Gly

 1
 5
 5
 10
 10
 10
 15
 15

 Ala
 Asp
 Glu
 Arg
 Ala
 Gly
 Pro
 Ala
 Arg
 Ile
 Try
 Arg
 Glu
 His

 Ser
 Met
 Ala
 Met
 Lys
 Pro
 Arg
 Thr
 Gly
 Asp
 Gly
 Pro
 Leu
 Glu
 Ala
 Ala

 Thr
 Lys
 Glu
 Gly
 Arg
 Gly
 Met
 Arg
 Val
 Pro
 Leu
 Glu
 Gly
 Arg
 Leu
 Gly
 Arg
 Gly
 Arg
 Fro
 Arg
 Fro
 Arg
 Gly
 Arg
 Gly
 Arg
 Fro
 Arg
 Fro
 Arg
 Fro
 Arg
 Fro
 Arg
 Fro
 Fro
 Arg
 Fro
 Fro
 Arg
 Fro
 Fro

(2) INFORMATION FOR SEQ ID NO:88:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 95 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

WO 99/42076 PCT/US99/03268

114

85 90 95

(2) INFORMATION FOR SEQ ID NO:89:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 166 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Met Thr Gln Ser Gln Thr Val Thr Val Asp Gln Gln Glu Ile Leu Asn

1 10 16

Arg Ala Asn Glu Val Glu Ala Pro Met Ala Asp Pro Pro Thr Asp Val

Pro Ile Thr Pro Cys Glu Leu Thr Xaa Xaa Lys Asn Ala Ala Gln Gln
35 40 45

Xaa Val Leu Ser Ala Asp Asn Met Arg Glu Tyr Leu Ala Ala Gly Ala
50 55 60

Lys Glu Arg Gln Arg Leu Ala Thr Ser Leu Arg Asn Ala Ala Lys Xaa 65 70 75 80

Tyr Gly Glu Val Asp Glu Glu Ala Ala Thr Ala Leu Asp Asn Asp Gly

Glu Gly Thr Val Gln Ala Glu Ser Ala Gly Ala Val Gly Gly Asp Ser

Ser Ala Glu Leu Thr Asp Thr Pro Arg Val Ala Thr Ala Gly Glu Pro

Asn Phe Met Asp Leu Lys Glu Ala Ala Arg Lys Leu Glu Thr Gly Asp

Gln Gly Ala Ser Leu Ala His Xaa Gly Asp Gly Trp Asn Thr Xaa Thr 145 150 155 160

Leu Thr Leu Gln Gly Asp

(2) INFORMATION FOR SEQ ID NO:90:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

Arg Ala Glu Arg Met

- (2) INFORMATION FOR SEQ ID NO:91:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 263 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

Val Ala Trp Met Ser Val Thr Ala Gly Gln Ala Glu Leu Thr Ala Ala 10 Gln Val Arg Val Ala Ala Ala Tyr Glu Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala Glu Asn Arg Ala Glu Leu Met Ile Leu 40 Ile Ala Thr Asn Leu Leu Gly Gln Asn Thr Pro Ala Ile Ala Val Asn Glu Ala Glu Tyr Gly Glu Met Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Ala Thr Ala Thr Ala Thr Ala Thr Leu Leu Pro Phe Glu Glu Ala Pro Glu Met Thr Ser Ala Gly Gly Leu Leu Glu Gln Ala 105 Ala Ala Val Glu Glu Ala Ser Asp Thr Ala Ala Ala Asn Gln Leu Met 120 Asn Asn Val Pro Gln Ala Leu Lys Gln Leu Ala Gln Pro Thr Gln Gly 135 Thr Thr Pro Ser Ser Lys Leu Gly Gly Leu Trp Lys Thr Val Ser Pro 150 His Arg Ser Pro Ile Ser Asn Met Val Ser Met Ala Asn Asn His Met 170 Ser Met Thr Asn Ser Gly Val Ser Met Thr Asn Thr Leu Ser Ser Met 185 Leu Lys Gly Phe Ala Pro Ala Ala Ala Ala Gln Ala Val Gln Thr Ala 195 200 Ala Gln Asn Gly Val Arg Ala Met Ser Ser Leu Gly Ser Ser Leu Gly 215 220 Ser Ser Gly Leu Gly Gly Gly Val Ala Ala Asn Leu Gly Arg Ala Ala 230 235 Ser Val Arg Tyr Gly His Arg Asp Gly Gly Lys Tyr Ala Xaa Ser Gly 245 250 Arg Arg Asn Gly Gly Pro Ala 260

(2) INFORMATION FOR SEQ ID NO:92:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 303 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

55 Glu Leu Gly Gly Gly Asp Gly Ala Val Ser Gly Asp Thr Gly Leu Pro 70 75 Val Gly Val Ala Leu Leu Ala Ala Leu Leu Ala Gly Val Val Leu Val 85 90 Pro Lys Ala Lys Ser His Val Thr Val Val Ala Val Leu Gly Val Leu 105 Gly Val Phe Leu Met Val Ser Ala Thr Phe Asn Lys Pro Ser Ala Tyr 120 Ser Thr Gly Trp Ala Leu Trp Val Val Leu Ala Phe Ile Val Phe Gln 135 Ala Val Ala Ala Val Leu Ala Leu Leu Val Glu Thr Gly Ala Ile Thr 150 Ala Pro Ala Pro Arg Pro Lys Phe Asp Pro Tyr Gly Gln Tyr Gly Arg 170 Tyr Gly Gln Tyr Gly Gln Tyr Gly Val Gln Pro Gly Gly Tyr Tyr Gly 185 Gln Gln Gly Ala Gln Gln Ala Ala Gly Leu Gln Ser Pro Gly Pro Gln 200 Gln Ser Pro Gln Pro Pro Gly Tyr Gly Ser Gln Tyr Gly Gly Tyr Ser 215 Ser Ser Pro Ser Gln Ser Gly Ser Gly Tyr Thr Ala Gln Pro Pro Ala 230 235 Gln Pro Pro Ala Gln Ser Gly Ser Gln Gln Ser His Gln Gly Pro Ser 245 250 Thr Pro Pro Thr Gly Phe Pro Ser Phe Ser Pro Pro Pro Pro Val Ser 265 Ala Gly Thr Gly Ser Gln Ala Gly Ser Ala Pro Val Asn Tyr Ser Asn 280 Pro Ser Gly Gly Glu Gln Ser Ser Pro Gly Gly Ala Pro Val 295

(2) INFORMATION FOR SEQ ID NO:93:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

Gly Cys Gly Glu Thr Asp Ala Ala Thr Leu Ala Gln Glu Ala Gly Asn 1 5 5 1 10 10 15 Phe Glu Arg Ile Ser Gly Asp Leu Lys Thr Gln Ile

(2) INFORMATION FOR SEQ ID NO:94:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 16 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

Asp Gln Val Glu Ser Thr Ala Gly Ser Leu Gln Gly Gln Trp Arg Gly 1 5 10 15

- (2) INFORMATION FOR SEQ ID NO:95:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

- (2) INFORMATION FOR SEQ ID NO:96:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

Gly Cys Gly Gly Thr Ala Ala Gln Ala Ala Val Val Arg Phe Gln Glu

1 5 5 1 10 15

Ala Ala Asn Lys Gln Lys Gln Glu Leu Asp Glu

20 25

- (2) INFORMATION FOR SEQ ID NO:97:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

Gly Cys Gly Ala Asn Lys Gln Lys Gln Glu Leu Asp Glu Ile Ser Thr 1 5 10 10 15 Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg

- (2) INFORMATION FOR SEQ ID NO:98:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 amino acids

- (B) TYPE: amino acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

Gly Cys Gly Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg Ala Asp Glu 1 5 10 15 15 Glu Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe 20 25

(2) INFORMATION FOR SEQ ID NO:99:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 507 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

ATGAAGATGG GCGGCCGGTG GTCTTCGGCG CAGTTGACCA AGTCTGGTCG AAGGCCGCCG GCCGCCGGTT ACGCAGAACG	TGACTTCGAT CGCCACTGCC GCCTGCTCAA AGGGCGGCAT AGCACGGGGA CGGCCACCGC TCACGTTCGT	CATGGCTGGC GTTGGACCCG CAGCCTCGCC CGGGGGCACC TCTGCCGCTG CGACGTTTCC GAATCAAGGC	GGCCCGGTCG GCATCCGCCC GATCCCAACG GAGGCGCGCA TCGTTCAGCG GTCTCGGGTC	TATACCAGAT CTGACGTCCC TGTCGTTTGC TCGCCGACCA TGACGAACAT CGAAGCTCTC	GCAGCCGGTC GACCGCCGCC GAACAAGGGC CAAGCTGAAG CCAGCCGGCG	60 120 180 240 300 360 420 480
GAGTTGCTGC	AGGCCGCAGG	GAACTGA	GGCTGGATGC	TGTCACGCGC	ATCGGCGATG	480 507

(2) INFORMATION FOR SEQ ID NO:100:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 168 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

 Met
 Lys
 Val
 Lys
 See
 Ile
 Ala
 Ala
 Gly
 Leu
 Thr
 Ala
 Ala
 Ala
 Ala

 Ile
 Gly
 Ala
 A

Ser	Val	Thr	100 Asn	Ile	Gln	Pro	Ala	105 Ala	Ala	Gly	Ser	Ala	110 Thr	Ala	Asp
		TT2	Ser				120				Val	125			_
Thr 145	Phe	Val	Asn	Gln	Gly 150		Trp	Met	Leu	Ser	140 Arg	Ala	Ser	Ala	Met 160
Glu	Leu	Leu	Gln	Ala 165	Ala	Gly	Asn								190

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 500 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

CGTGGCAATG CATTAACACC GGCTGCCGCA CGCACCGCCA ACAGTACATC GGGCCCCATC ACGGGCCGCA CAACGGGCCGCA CAACGGGCCGCA	ACCTGCAATT CAGTTCAACG CCTCAGCGCG GGCCTTGTCG CCGCGACCCG TCCCGCGACC CATCTCGTGC	ACGGGCAGGT CCTCACCGGT CTGCCATGGC AGTCGGTTGC GCATCGTCGC CGGCATCGTC	AGTAGCTGCG GGCGCAGTCC CGCGCAATTG CGGCTCCTGC CGGGGCTAGG GCCGGGGCTA	CTCAACGCGA TATTTGCGCA CAAGCTGTGC AACAACTATT CCAGATTGCC GGCCAGATTG	CGGATCCGGG ATTTCCTCGC CGGGGGCGGC AAGCCCATGC CCGCTCCTCA	60 120 180 240 300 360 420 480
GCCGCCACCG	CGGTGGAGCT	committee	CAGCCCGGGG	GATCCACTAG	TTCTAGAGCG	480 500

(2) INFORMATION FOR SEQ ID NO:102:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 96 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

 Val
 Ala
 Met
 Ser
 Leu
 Thr
 Val
 Gly
 Ala
 Gly
 Val
 Ala
 Ser
 Ala
 Asp
 Pro

 Val
 Asp
 Ala
 Tle
 Asp
 Thr
 Thr
 Cys
 Asp
 Tyr
 Gly
 G

(2) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 154 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:	
ATGACAGAGC AGCAGTGGAA TTTCGCGGGT ATCGAGGCCG CGGCAAGCGC AATCCAGGGA AATGTCACGT CCATTCATTC CCTCCTTGAC GAGGGGAAGC AGTCCCTGAC CAAGCTCGCA GCGGCCTGGG GCGGTAGCGG TTCGGAAGCG TACC	60 120 154
(2) INFORMATION FOR SEQ ID NO:104:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 51 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:	
Met Thr Glu Gln Gln Trp Asn Phe Ala Gly Ile Glu Ala Ala Ala Ser 1 5 10 15 Ala Ile Gln Gly Asn Val Thr Ser Ile His Ser Leu Leu Asp Glu Gly	
20 25 30 Lys Gln Ser Leu Thr Lys Leu Ala Ala Ala Trp Gly Gly Ser Gly Ser 35 40 45 Glu Ala Tyr 50	
(2) INFORMATION FOR SEQ ID NO:105:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 282 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:	
GGTCGCGCA CTTCCAGGTG ACTATGAAAG TCGGCTTCCG NCTGGAGGAT TCCTGAACCT	60
CTCGAAACG CGGCACAGCC GACGGTGGCT CCGNCGAGGC GCTGNCTCCA AAATCCCTGA ACAATTCGN CGGGGGCGCC TACAAGGAAG TCGGTGCTGA ATTCGNCGNG TATCTGGTCG ACCTGTGTGG TCTGNAGCCG GACGAAGCGG TGCTCGACGT CG	120 180 240 282
2) INFORMATION FOR SEQ ID NO:106:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 3058 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single 	

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

GATCGTACCC GTGCGAGTGC TCGGGCCGTT TGAGGATGGA GTGCACGTGT CTTTCGTG	AT 60
GGCATACCCA GAGATGTTGG CGGCGGCGGC TGACACCCTG CAGAGCATCG GTGCTACC	AC 120
TGTGGCTAGC AATGCCGCTG CGGCGGCCCC GACGACTGGG GTGGTGCCCC CCGCTGCC	GA 180
TGAGGTGTCG GCGCTGACTG CGGCGCACTT CGCCGCACAT GCGGCGATGT ATCAGTCC	GT 240
GAGCGCTCGG GCTGCTGCGA TTCATGACCA GTTCGTGGCC ACCCTTGCCA GCAGCGCC	'AG 300
CTCGTATGCG GCCACTGAAG TCGCCAATGC GGCGGCGGCC AGCTAAGCCA GGAACAGT	CG 360
GCACGAGAAA CCACGAGAAA TAGGGACACG TAATGGTGGA TTTCGGGGCG TTACCACC	GG 420
AGATCAACTC CGCGAGGATG TACGCCGGCC CGGGTTCGGC CTCGCTGGTG GCCGCGGC	TC 480
AGATGTGGGA CAGCGTGGCG AGTGACCTGT TTTCGGCCGC GTCGGCGTTT CAGTCGGT	
TCTGGGGTCT GACGGTGGGG TCGTGGATAG GTTCGTCGGC GGGTCTGATG GTGGCGGC	
CCTCGCCGTA TGTGGCGTGG ATGAGCGTCA CCGCGGGGCA GGCCGAGCTG ACCGCCGC	
AGGTCCGGGT TGCTGCGGCG GCCTACGAGA CGGCGTATGG GCTGACGGTG CCCCCGCC	
TGATCGCCGA GAACCGTGCT GAACTGATGA TTCTGATAGC GACCAACCTC TTGGGGCA	
ACACCCCGGC GATCGCGGTC AACGAGGCCG AATACGGCGA GATGTGGGCC CAAGACGC	
CCGCGATGTT TGGCTACGCC GCGGCGACGG CGACGGCGAC GGCGACGTTG CTGCCGTTC	
AGGAGGCGCC GGAGATGACC AGCGCGGGTG GGCTCCTCGA GCAGGCCGCC GCGGTCGAC	
AGGCCTCCGA CACCGCCGCG GCGAACCAGT TGATGAACAA TGTGCCCCAG GCGCTGCAI	
AGCTGGCCCA GCCCACGCAG GGCACCACGC CTTCTTCCAA GCTGGGTGGC CTGTGGAAC	
CGGTCTCGCC GCATCGGTCG CCGATCAGCA ACATGGTGTC GATGGCCAAC AACCACATC	
CGATGACCAA CTCGGGTGTG TCGATGACCA ACACCTTGAG CTCGATGTTG AAGGGCTTT	
CTCCGGCGGC GGCCGCCAG GCCGTGCAAA CCGCGGCGCA AAACGGGGTC CGGGCGATC	
GCTCGCTGGG CAGCTCGCTG GGTTCTTCGG GTCTGGGCGG TGGGGTGGCC GCCAACTTC	
GTCGGGCGGC CTCGGTCGGT TCGTTGTCGG TGCCGCAGGC CTGGGCCGCG GCCAACCAG	
CAGTCACCCC GGCGGCGCGG GCGCTGCCGC TGACCAGCCT GACCAGCGCC GCGGAAAGA	
GGCCCGGGCA GATGCTGGGC GGGCTGCCGG TGGGGCAGAT GGGCGCCAGG GCCGGTGGT	
GGCTCAGTGG TGTGCTGCGT GTTCCGCCGC GACCCTATGT GATGCCGCAT TCTCCGGCG	
CCGGCTAGGA GAGGGGGCGC AGACTGTCGT TATTTGACCA GTGATCGGCG GTCTCGGTG	T 1620

TTCCCCCCCC					
TTCCGCGGCC GGCTATGAC					1680
GTTCAACAAG GAGACAGGC					1740
GGACATGGCG GGCCGTTTT	G AGGTGCACG	CCAGACGGT	GAGGACGAG	CTCGCCGGAT	1800
GTGGGCGTCC GCGCAAAACI					1860
GCTAGACACC ATGGCCCAG					1920
GGTGCGTGAC GGGCTGGTTC					
GCAGATCCTC AGCAGCTAAC					1980
ACAGGTTCGA TGACCATCAA					2040
CGCGCTCAGG CCGGGTTGCT					2100
					2160
GCGAGTGACT TTTGGGGCGG					2220
GGCCGTAACT TCCAGGTGAT					2280
GCCGGCAACA ACATGGCGCA					2340
GCCAAGGCCA GGGACGTGGT	GTACGAGTGA	AGTTCCTCGC	GTGATCCTTC	GGGTGGCAGT	2400
CTAAGTGGTC AGTGCTGGGG	TGTTGGTGGT	TTGCTGCTTG	GCGGGTTCTT	CGGTGCTGGT	2460
CAGTGCTGCT CGGGCTCGGG					2 52 0
TTCGTCGTGT TGTTCGGCGA					2580
GAAGATGCCC ACGACGTCGG					
GTTGGACCAG ATTTGGCGCC					2640
GCGGGCGGTG TCGAGGTGCT					2700
CCGATCATAT TGGGCAACAA					2760
					2820
GGTGCGCACC CACGGCCAGG					2880
GGTTCTGCAG CGCTGCCAGG					2940
GGCGTGGGCG TCGCTGGTGA					3000
GAAGAACGCC AGCCAGCCGG	CCCGTCCTC	GGCGGAGGTG :	ACCTGGATGC (CCAGGATC	3058

(2) INFORMATION FOR SEQ ID NO:107:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 391 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single

WO 99/42076 PCT/US99/03268

123

(D) TOPOLOGY: linear

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:
- Met Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met

 1 5 10 15
- Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Gln Met Trp
 20 25 30
- Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser
- Val Val Trp Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly 50 55 60
- Leu Met Val Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr 75 75 80
- Ala Gly Gln Ala Glu Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala 85 90 95
- Ala Tyr Glu Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala
- Glu Asn Arg Ala Glu Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Gly
- Gln Asn Thr Pro Ala Ile Ala Val Asn Glu Ala Glu Tyr Gly Glu Met 130 135 140
- Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Ala Thr Ala 145 150 155 160
- Thr Ala Thr Ala Thr Leu Leu Pro Phe Glu Glu Ala Pro Glu Met Thr 165 170 175
- Ser Ala Gly Gly Leu Leu Glu Gln Ala Ala Ala Val Glu Glu Ala Ser 180 185 190
- Asp Thr Ala Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu
 195 200 205
- Gln Gln Leu Ala Gln Pro Thr Gln Gly Thr Thr Pro Ser Ser Lys Leu 210 215 220
- Gly Gly Leu Trp Lys Thr Val Ser Pro His Arg Ser Pro Ile Ser Asn 225 230 235 240
- Met Val Ser Met Ala Asn Asn His Met Ser Met Thr Asn Ser Gly Val
- Ser Met Thr Asn Thr Leu Ser Ser Met Leu Lys Gly Phe Ala Pro Ala 260 265 270

Ala	Ala	1 Ala 275	Glr	Ala	Val	Gln	Thr 280	Ala	Ala	Gln	Asn	Gly 285		Arg	Ala
Met	Ser 290	Ser	Leu	Gly	Ser	Ser 295	Leu	Gly	Ser	Ser	Gly 300	Leu	Gly	Gly	Gly
Val 305	Ala	Ala	Asn	Leu	Gly 310	Arg	Ala	Ala	Ser	Val 315	Gly	Ser	Leu	Ser	Val 320
Pro	Gln	Ala	Trp	Ala 325	Ala	Ala	Asn	Gln	Ala 330	Val	Thr	Pro	Ala	Ala 335	Arg
Ala	Leu	Pro	Leu 340	Thr	Ser	Leu	Thr	Ser 345	Ala	Ala	Glu	Arg	Gly 350	Pro	Gly
Gln	Met	Leu 355	Gly	Gly	Leu	Pro	Val 360	Gly	Gln	Met	Gly	Ala 365	Arg	Ala	Gly
Gly	Gly 370	Leu	Ser	Gly	Val	Leu 375	Arg	Val	Pro	Pro	Arg 380	Pro	Tyr	Val	Met
Pro 385	His	Ser	Pro	Ala .	Ala 390	Gly									

(2) INFORMATION FOR SEQ ID NO:108:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1725 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

	CNCCMCNCC						
						GGTCAAGGTG	60
	ACGTCCCTCG	GCGTGTCGCC	GGCGTGGATG	CAGACTCGAT	GCCGCTCTTT	AGTGCAACTA	120
	ATTTCGTTGA	AGTGCCTGCG	AGGTATAGGA	CTTCACGATT	GGTTAATGTA	GCGTTCACCC	180
				GTCACCAACG			
				CTTGAGCTCC			240
							300
				GCAACCGGAG			360
			CAACGCTGAA	GGCTAGCAAT	GCCGCCGCAG	CCGTGCCGAC	420
(GACTGGGGTG	GTGCCCCCGG	CTGCCGACGA*	GGTGTCGCTG	CTGCTTGCCA	CACAATTCCG	480
•	FACGCATGCG	GCGACGTATC	AGACGGCCAG	CGCCAAGGCC	GCGGTGATCC	ATGAGCAGTT	540
				ATATGCGGAC			
				GAGCGGAAGG			600
			_		MATATCGAAG	TGGTGGATTT	660

CCCCCCCC						
					GTTCGGCCTC	720
GCTGGTGGC	C GCCGCGAAG	A TGTGGGACAC	G CGTGGCGAG	r GACCTGTTT	CGGCCGCGTC	780
GGCGTTTCA	G TCGGTGGTCT	GGGGTCTGAC	GGTGGGGTC	G TGGATAGGTT	CGTCGGCGGG	840
TCTGATGGC	G GCGGCGGCCT	CGCCGTATGT	GGCGTGGATC	G AGCGTCACCG	CGGGGCAGGC	900
CCAGCTGAC	C GCCGCCCAGG	TCCGGGTTGC	TGCGGCGGCC	TACGAGACAG	CGTATAGGCT	960
GACGGTGCC	C CCGCCGGTGA	TCGCCGAGAA	CCGTACCGAA	CTGATGACGC	TGACCGCGAC	1020
CAACCTCTTC	GGGCAAAACA	CGCCGGCGAT	CGAGGCCAAT	CAGGCCGCAT	ACAGCCAGAT	1080
	GACGCGGAGG					1140
	CCGTTCGAGG					1200
	GTCGAGGAGG					1260
	CTGCAACAGC					1320
	TGGACGGCGG					1380
	CACATGTCGA					1440
	GGCTTAGCTC					1500
	ATGAGCTCGC					1560
	GTGGCCGCCA					1620
	GCCGCGGCCA .					1680
	AGCGCCGCCC					1725

(2) INFORMATION FOR SEQ ID NO:109:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 359 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

Val Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met
1 5 10 15

Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Lys Met Trp

Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser

							25	,					60					Gly
Le 65	u M	et A	la	Ala	Al	a Al 70	a Se	r Pr	0	Tyr	Va.	l Al 75	a Tr	p M e	t Se	er V	al	Thr 80
Al	a G	ly G	ln	Ala	Gl: 85	n Le	u Th	r Al	а.	Ala	Glr 90	ı Va	l Ar	g Va	1 A1	.a A. 95		Ala
Al	a Ty	Æ G	lu	Thr 100	Ala	а Ту:	r Ar	g Le	u !	Thr 105	Va]	. Pr	o Pr	o Pr	o Va		Le	Ala
Gl	u As	n A	rg 15	Thr	Glu	Le:	ı Me	t Th	r 1	Leu	Thr	Ala	a Th	As:		u Le	ıu	Gly
Gli	n As 13	n Ti O	hr 1	Pro	Ala	Ile	Gl:	u Ala	a 2	Asn	Gln	Ala	140	i Ту:	r Se	r Gl	n i	Met
T::	Gl 5	y G]	ln 1	qzA	Ala	Glu 150	ı Ala	a Met	- 1	lyr	Gly	Tyr 155	: Ala	Ala	a Th	r Al		Ala 160
Thr	Al.	a Th	r c	Slu	Ala 165	Leu	Let	Pro	P	he	Glu 170	Asp	Ala	Pro) Le	u Il 17		Thr
Asn	Pro	Gl	у G 1	1y .80	Leu	Leu	Glu	Gln	1 A	la .85	Val	Ala	Val	Glu	Glu 190		a)	Ile
Asp	Thi	Al 19	a A 5	la.	Ala	Asn	Gln	Leu 200	M	et	Asn	Asn	Val	Pro	Glr	ı Ala	a I	Leu
Gln	Glr 210	Le	u A	la (Gln	Pro	Ala 215	Gln	. G.	ly	Val	Val	Pro 220	Ser	Ser	: Lys	3 1	-eu
Gly 225	Gly	Lei	u T:	ָ מַז	Thr	Ala 230	Val	Ser	P:	ro:	His	Leu 235	Ser	Pro	Leu	Ser		sn 40
Val	Ser	Ser	: I	le A	Ala 245	Asn	Asn	His	Me	et s	Ser 250	Met	Met	Gly	Thr	Gly 255		al
Ser	Met	Thr	26	sn 1	Thr	Leu	His	Ser	Ме 2 е	et 1	Leu	Lys	Gly	Leu	Ala 270	Pro	A	la
Ala	Ala	Gln 275	Al	.a v	al ·	Glu	Thr	Ala 280	Al	la c	Slu	Asn	Gly	Val 285	Trp	Ala	M	et
Ser	Ser 290	Leu	. Gl	y S	er (Gln	Leu 295	Gly	Se	er s	Ser	Leu	Gly 300	Ser	Ser	Gly	L	eu
Gly 305					•							315					32	20
Leu	Ser	Val	Pr	o P: 3:	ro 2 25	Ala	Trp	Ala	Al	a A 3	la 2 30	lsn.	Gln	Ala	Val	Thr	Pı	ro

Ala Ala Arg Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Gln Thr 340 345 350

Ala Pro Gly His Met Leu Gly 355

(2) INFORMATION FOR SEQ ID NO:110:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3027 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

AGTTCAGTCG AGAATGATAC TGACGGGCTG TATCCACGAT GGCTGAGACA ACCGAACCAC 60 CGTCGGACGC GGGGACATCG CAAGCCGACG CGATGGCGTT GGCCGCCGAA GCCGAAGCCG 120 CCGAAGCCGA AGCGCTGGCC GCCGCGGCGC GGGCCCGTGC CCGTGCCGCC CGGTTGAAGC 180 GTGAGGCGCT GGCGATGGCC CCAGCCGAGG ACGAGAACGT CCCCGAGGAT ATGCAGACTG 240 GGAAGACGCC GAAGACTATG ACGACTATGA CGACTATGAG GCCGCAGACC AGGAGGCCGC 300 ACGGTCGGCA TCCTGGCGAC GGCGGTTGCG GGTGCGGTTA CCAAGACTGT CCACGATTGC 360 CATGGCGGCC GCAGTCGTCA TCATCTGCGG CTTCACCGGG CTCAGCGGAT ACATTGTGTG 420 GCAACACCAT GAGGCCACCG AACGCCAGCA GCGCGCCGCG GCGTTCGCCG CCGGAGCCAA 480 GCAAGGTGTC ATCAACATGA CCTCGCTGGA CTTCAACAAG GCCAAAGAAG ACGTCGCGCG 540 TGTGATCGAC AGCTCCACCG GCGAATTCAG GGATGACTTC CAGCAGCGGG CAGCCGATTT 600 CACCAAGGTT GTCGAACAGT CCAAAGTGGT CACCGAAGGC ACGGTGAACG CGACAGCCGT 660 CGAATCCATG AACGAGCATT CCGCCGTGGT GCTCGTCGCG GCGACTTCAC GGGTCACCAA 720 TTCCGCTGGG GCGAAAGACG AACCACGTGC GTGGCGGCTC AAAGTGACCG TGACCGAAGA 780 GGGGGGACAG TACAAGATGT CGAAAGTTGA GTTCGTACCG TGACCGATGA CGTACGCGAC 840 GTCAACACCG AAACCACTGA CGCCACCGAA GTCGCTGAGA TCGACTCAGC CGCAGGCGAA 900 GCCGGTGATT CGGCGACCGA GGCATTTGAC ACCGACTCTG CAACGGAATC TACCGCGCAG 960 AAGGGTCAGC GGCACCGTGA CCTGTGGCGA ATGCAGGTTA CCTTGAAACC CGTTCCGGTG 1020 ATTCTCATCC TGCTCATGTT GATCTCTGGG GGCGCGACGG GATGGCTATA CCTTGAGCAA 1080 TACGACCCGA TCAGCAGACG GACTCCGGCG CCGCCCGTGC TGCCGTCGCC GCGGCGTCTG 1140 ACGGGACAAT CGCGCTGTTG TGTATTCACC CGACACGTCG ACCAAGACTT CGCTACCGCC 1200

AGGTCGCACC TCGCCGGCGA TTTCCTGTCC TATACGACCA GTTCACGCAG CAGATCGTGG	1250
CTCCGGCGGC CAAACAGAAG TCACTGAAAA CCACCGCCAA GGTGGTGCGC GCGGCCGTGT	1260
	1320
CGGAGCTACA TCCGGATTCG GCCGTCGTTC TGGTTTTTGT CGACCAGAGC ACTACCAGTA	1380
AGGACAGCCC CAATCCGTCG ATGGCGGCCA GCAGCGTGAT GGTGACCCTA GCCAAGGTCG	1440
ACGGCAATTG GCTGATCACC AAGTTCACCC CGGTTTAGGT TGCCGTAGGC GGTCGCCAAG	1500
TCTGACGGGG GCGCGGGTGG CTGCTCGTGC GAGATACCGG CCGTTCTCCG GACAATCACG	1560
GCCCGACCTC AAACAGATCT CGGCCGCTGT CTAATCGGCC GGGTTATTTA AGATTAGTTG	1620
CCACTGTATT TACCTGATGT TCAGATTGTT CAGCTGGATT TAGCTTCGCG GCAGGGCGGC	1680
TGGTGCACTT TGCATCTGGG GTTGTGACTA CTTGAGAGAA TTTGACCTGT TGCCGACGTT	1740
GTTTGCTGTC CATCATTGGT GCTAGTTATG GCCGAGCGGA AGGATTATCG AAGTGGTGGA	1800
CTTCGGGGCG TTACCACCGG AGATCAACTC CGCGAGGATG TACGCCGGCC CGGGTTCGGC	1860
CTCGCTGGTG GCCGCCGCGA AGATGTGGGA CAGCGTGGCG AGTGACCTGT TTTCGGCCGC	1920
GTCGGCGTTT CAGTCGGTGG TCTGGGGTCT GACGACGGGA TCGTGGATAG GTTCGTCGGC	
GGGTCTGATG GTGGCGGCGG CCTCGCCGTA TGTGGCGTGG ATGAGCGTCA CCGCGGGGCA	1980
GGCCGAGCTG ACCGCCGCCC AGGTCCGGGT TGCTGCGGCG GCCTACGAGA CGGCGTATGG	2040
GCTGACGGTG CCCCCCCCC TONNESS OF TGCTGCGGCG GCCTACGAGA CGGCGTATGG	2100
GCTGACGGTG CCCCCGCCGG TGATCGCCGA GAACCGTGCT GAACTGATGA TTCTGATAGC	2160
GACCAACCTC TTGGGGCAAA ACACCCCGGC GATCGCGGTC AACGAGGCCG AATACGGGGA	2220
GATGTGGGCC CAAGACGCCG CCGCGATGTT TGGCTACGCC GCCACGGCGG CGACGGCGAC	2280
CGAGGCGTTG CTGCCGTTCG AGGACGCCCC ACTGATCACC AACCCCGGCG GGCTCCTTGA	2340
GCAGGCCGTC GCGGTCGAGG AGGCCATCGA CACCGCCGCG GCGAACCAGT TGATGAACAA	2400
TGTGCCCCAA GCGCTGCAAC AACTGGCCCA GCCCACGAAA AGCATCTGGC CGTTCGACCA	2460
ACTGAGTGAA CTCTGGAAAG CCATCTCGCC GCATCTGTCG CCGCTCAGCA ACATCGTGTC	2520
GATGCTCAAC AACCACGTGT CGATGACCAA CTCGGGTGTG TCGATGGCCA GCACCTTGCA	2580
CTCAATGTTG AAGGGCTTTG CTCCGGCGGC GGCTCAGGCC GTGGAAACCG CGGCGCAAAA	2640
CGGGGTCCAG GCGATGAGCT CGCTGGGCAG CCAGCTGGGT TCGTCGCTGG GTTCTTCGGG	2700
TCTGGGCGCT GGGGTGGCCG CCAACTTGGG TCGGGCGGCC TCGGTCGGTT CGTTGTCGGT	
GCCGCAGGCC TGGGCCGCGG CCAACCAGGC GGTCACCCCG GCGGCGCGGG CGCTGCCGCT	2760

WO 99/42076 PCT/US99/03268

GACCAG	CCTG	ACC	'AGCG	CCG	CCCA	AACC	GC C	CCCG	SACA	C ATO	CTG	GCG	GGCT	ACCG	cr
GGGGCA															
GCCGCG															
TGCGGG															
(2) IN	FORM	ATIO	N FOI	R SE() ID	NO:1	.11:								
	((A) I (B) I (C) S (D) I	ENGT TYPE: TRAN	TH: 3 ami DEDN OGY:	no a ESS:	ear	aci								
						ON:									
				_					10					15	Met
								45					30		Trp
							40					45	Phe		
Va]	L Val 50	l Tr	Gly	/ Let	ı Thi	Thr 55	Gly	'Ser	Trp	Ile	Gly 60	Ser	Ser	Ala	Gly
Le u 65	Met	: 'Yal	. Ala	Ala	70	Ser	Pro	Tyr	Val	Ala 75	Trp	Met	Ser	Val	Thr 80
									90				Ala	95	
								105					Val		
							120					125	Leu		
						+33					140		Gly		
Trp 145	Ala	Gln	Asp	Ala	Ala 150	Ala	Met	Phe	Gly	Tyr 155	Ala	Ala	Thr		Ala 160
Thr	Ala	Thr	Glu	Ala 165	Leu	Leu⁼	Pro	Phe	Glu 170	Asp	Ala	Pro	Leu	Ile 175	Thr
Asn	Pro	Gly	Gly 180	Leu	Leu	Glu	Gln	Ala 185	Val	Ala	Val	Glu	Glu 190	Ala	Ile
Ąsp	Thr	Ala	Ala	Ala	Asn	Gln	Leu	Met	Asn .	Asn	Val	Pro	Gln .	Ala	Leu

			19	5				20	0				205	;			
	Gli	n Gl: 21	n Lei	u Ala	Glr	Pro	Thr 215	Lys	s Ser	Ile	: Trp	Pro 220	Phe	Asp	Gln	Leu	
	Ser 225	Glu	ı Lei	ı Trp	Lys	Ala 230	Ile	e Ser	Pro	His	Leu 235	Ser	Pro	Leu	Ser	Asn 240	
	Ile	. Val	. Ser	Met	Leu 245	Asn	Asn	His	Val	Ser 250	Met	Thr	Asn	Ser	Gly 255		
	Ser	Met	Ala	Ser 260	Thr	Leu	His	Ser	Me t 265	Leu	Lys	Gly	Phe	Ala 270	Pro	Ala	
	Ala	Ala	Gln 275	Ala	Val	Glu	Thr	Ala 280	Ala	Gln	Asn	Gly	Val 285	Gln	Ala	Met	
	Ser	Ser 290	Leu	Gly	Ser	Gln	Leu 295	Gly	Ser	Ser	Leu	Gly 300	Ser	Ser	Gly	Leu	
	Gly 30 5	Ala	Gly	Val	Ala	Ala 310	Asn	Leu	Gly	Arg	Ala 315	Ala	Ser	Val	Gly	Ser 320	
	Leu	Ser	Val	Pro	Gln 325	Ala	Trp	Ala	Ala	Ala 330	Asn	Gln	Ala	Val	Thr 335	Pro	
,	Ala	Ala	Arg	Ala 340	Leu	Pro	Leu	Thr	Ser 345	Leu	Thr	Ser	Ala	Ala 350	Gln	Thr	
i	Ala	Pro	Gly 355	His	Met	Leu	Gly	Gly 360	Leu	Pro	Leu		Gln 365	Leu	Thr	Asn	
٤	Ser	Gly 370	Gly	Gly	Phe (Gly (Gly 375	Val	Ser .	Asn		Leu . 380	Arg	Met	Pro	Pro	
, 3	rg :	Ala	Tyr	Val 1	Met :	Pro <i>1</i> 390	Arg	Val	Pro i		Ala (Gly					
(2) IN	FOR	MATI	ON F	OR SI	EQ II	ОИ С	:112	:									
(i) S	(A) (B) (C)	TYPI STR	STH: S: nu ANDEI	1616 Iclei NESS	ERIST bas ic ac 5: si near	se pa cid .ngle	airs									
(x.	i) s	EQUE	ENCE	DESC	RIPT	'ION :	SEC) ID	NO : 1	.12:							
CATCGG											:CGGA	GTAA	ATE	יככפי	'ACG		<i>c</i> ^
GCTGAT																	60
TTCGGCC																	120
AGCCTGG																	240

GCTACAAACC COOKE	
GCTACAAACC GCGTCAACAC AGGCCAAGAC CCGTGCGATG CAGGCGACGG CGCAAGCCGC	300
GGCATACACC CAGGCCATGG CCACGACGCC GTCGCTGCCG GAGATCGCCG CCAACCACAT	360
CACCCAGGCC GTCCTTACGG CCACCAACTT CTTCGGTATC AACACGATCC CGATCGCGTT	420
GACCGAGATG GATTATTTCA TCCGTATGTG GAACCAGGCA GCCCTGGCAA TGGAGGTCTA	
CCAGGCCGAG ACCGCGGTTA ACACGCTTTT CGAGAAGCTC GAGCCGATGG CGTCGATCCT	480
TGATCCCGGC GCGAGCCAGA GCACGACGAA CCCGATCTTC GGAATGCCCT CCCCTGGCAG	540
CTCAACACCG GTTGGCCAGT TGGGGGGGG	600
CTCAACACCG GTTGGCCAGT TGCCGCCGGC GGCTACCCAG ACCCTCGGCC AACTGGGTGA	660
GATGAGCGGC CCGATGCAGC AGCTGACCCA GCCGCTGCAG CAGGTGACGT CGTTGTTCAG	720
CCAGGTGGGC GGCACCGGCG GCGGCAACCC AGCCGACGAG GAAGCCGCGC AGATGGGCCT	780
GCTCGGCACC AGTCCGCTGT CGAACCATCC GCTGGCTGGT GGATCAGGCC CCAGCGCGGG	840
CGCGGGCCTG CTGCGCGCGG AGTCGCTACC TGGCGCAGGT GGGTCGTTGA CCCGCACGCC	900
GCTGATGTCT CAGCTGATCG AAAAGCCGGT TGCCCCCTCG GTGATGCCGG CGGCTGCTGC	960
CGGATCGTCG GCGACGGGTG GCGCCGCTCC GGTGGGTGCG GGAGCGATGG GCCAGGGTGC	
GCAATCCGGC GGCTCCACCA GGCCGGGTCT GGTCGCGCCG GCACCGCTCG CGCAGGAGCG	1020
TGAAGAAGAC GACGAGGACG ACTGGGACGA AGAGGACGAC TGGTGAGCTC CCGTAATGAC	1080
AACAGACTTC CCGGCCACCC CCGGCCACCC	1140
AACAGACTTC CCGGCCACCC GGGCCGGAAG ACTTGCCAAC ATTTTGGCGA GGAAGGTAAA	1200
GAGAGAAAGT AGTCCAGCAT GGCAGAGATG AAGACCGATG CCGCTACCCT CGCGCAGGAG	1260
GCAGGTAATT TCGAGCGGAT CTCCGGCGAC CTGAAAACCC AGATCGACCA GGTGGAGTCG	1320
ACGGCAGGTT CGTTGCAGGG CCAGTGGCGC GGCGCGGGGGGGGGG	1380
GTGGTGCGCT TCCAAGAAGC AGCCAATAAG CAGAAGCAGG AACTCGACGA GATCTCGACG	1440
AATATTCGTC AGGCCGGCGT CCAATACTCG AGGGCCGACG AGGAGCAGCA GCAGGCGCTG	
TCCTCGCAAA TGGGCTTCTG ACCCGCTAAT ACGAAAAGAA ACGGAGCAAA AACATGACAG	1500
AGCAGCAGTG GAATTTCGCG GGTATCGAGG CCGCGGCAAG CGCAATCCAG GGAAAT	1560
CCGCGGCAAG CGCAATCCAG GGAAAT	1616

(2) INFORMATION FOR SEQ ID NO:113:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 432 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

	(I) S	EQUE	NCE	DESC	RIPT	ION:	SEQ	ID	NO:1	13:					
CTAGTO	GATG	GGA	.CCAT	GGC	CATT	TTCT	GC A	GTCT	CACT	G CC	TTCT	GTGT	TGA	CATT	rtg
GCACGC	CGGC	GGA	AACG	AAG (CACT	GGGG'	TC G	AAGA	ACGG	C TG	CGCT	GCCA	TAT	CGTC	:GG
AGCTTC	CATA	CCT	TCGT	GCG (GCCG(GAAG	AG C	TTGT	CGTA	TC	GCC	GCCA	TGA	CAAC	TC
TCAGAG	TGCG	CTC	AAAC	GTA 7	[AAA]	CACG	AG A	AAGGG	GCGA	AC	CGAC	GAA	GGT	GAAC	TC
GCCCGA:	rccc	GTG:	TTC	GCT A	\TTC1	TACGO	G A	ACTCO	GCG1	TGO	CCT	\TGC	GAAC	ATC	CA
GTGACG:	MGC	CTT	GGT	GA A	LGCCA	TTG	C TO	ACCG	GCTI	. ccc	TGAT	CGT	CCGC	GCCA	.GG
TCTGC	I GCG	CGTI	GTTC	'AG C	TCGG	TAGO	C GI	GGCG	TCCC	ATI	TTTC	CTG	GACA	CCCT	GG
TACGCCI															
(2) I NF	ORMA	TION	FOR	SEQ	ID	NO:1	14:								
(i	(QUEN A) L B) T C) S T O	ENGT: YPE : TRANI	H: 30 amin DEDNI	68 an no ao ESS:	mino cid sine	aci	ds							
(xi) SE	QUEN	E DI	ESCRI	PTIC	ON: S	SEQ :	ID NO	D:114	ł:					
				-					10					15	Met
		/ Ala	Gly 20	Pro	Ala	Pro	Met	Leu	Ala	Ala	Ala	Ala	Gly	Tro	- G1
Thr	Leu							23					30		
							40	Gln	Ala	Val	Glu	Leu 45	30 Thr	Ala	Arg
Leu	Asn 50	Ser	Leu	Gly	Glu	Ala 55	Trp	Gln Thr	Ala Gly	Val Gly	Glu Gly 60	Leu 45 Ser	30 Thr Asp	Ala Lys	
Leu Leu 65	Asn 50 Ala	Ser Ala	Leu Ala	Gly	Glu Pro 70	Ala 55 Met	Trp	Gln Thr	Ala Gly Trp	Val Gly Leu 75	Glu Gly 60	Leu 45 Ser Thr	Thr Asp	Ala Lys Ser	Arg Ala Thr
Leu Leu 65 Gln	Asn 50 Ala Ala	Ser Ala Lys	Leu Ala Thr	Gly Thr Arg 85	Glu Pro 70 Ala	Ala 55 Met	Trp Val Gln	Gln Thr Val	Ala Gly Trp Thr 90	Val Gly Leu 75 Ala	Glu Gly 60 Gln	Leu 45 Ser Thr	Thr Asp Ala	Lys Ser Ala	Arg Ala Thr 80
Leu 65 Gln Thr	Asn 50 Ala Ala Gln	Ser Ala Lys Ala	Leu Ala Thr Met 100	Gly Thr Arg 85 Ala	Pro 70 Ala	Ala 55 Met Met	Trp Val Gln Pro	Gln Thr Val Ala Ser	Gly Trp Thr 90 Leu	Val Gly Leu 75 Ala	Glu Gly 60 Gln Gln	Leu 45 Ser Thr Ala	Thr Asp Ala Ala Ala 110	Lys Ser Ala 95	Arg Ala Thr 80 Tyr Asn
Leu 65 Gln Thr	Asn 50 Ala Ala Gln Ile	Ser Ala Lys	Leu Ala Thr Met 100	Gly Thr Arg 85 Ala Ala	Glu Pro 70 Ala Thr	Ala 55 Met Met Thr	Trp Val Gln Pro Thr	Gln Thr Val Ala Ser 105 Ala	Ala Gly Trp Thr 90 Leu	Vall Gly Leu 75 Ala Pro	Glu Gly 60 Gln Glu	Leu 45 Ser Thr Ala Ile	Thr Asp Ala Ala Ala Gly	Ala Lys Ser Ala 95 Ala	Arg Ala Thr 80 Tyr Asn

WO 99/42076 PCT/US99/03268

133

Asn	Gln	Ala	Ala	Leu	Ala	Met	C1	Val	m			_			
145					150	HEL	GIU	vai	TYT	GIn	Ala	Glu	Thr	Ala	Val
					150					155					160

- Asn Thr Leu Phe Glu Lys Leu Glu Pro Met Ala Ser Ile Leu Asp Pro
- Gly Ala Ser Gln Ser Thr Thr Asn Pro Ile Phe Gly Met Pro Ser Pro
- Gly Ser Ser Thr Pro Val Gly Gln Leu Pro Pro Ala Ala Thr Gln Thr
- Leu Gly Gln Leu Gly Glu Met Ser Gly Pro Met Gln Gln Leu Thr Gln
 210 215 220
- Pro Leu Gln Gln Val Thr Ser Leu Phe Ser Gln Val Gly Gly Thr Gly 235 230 240
- Gly Gly Asn Pro Ala Asp Glu Glu Ala Ala Gln Met Gly Leu Leu Gly 245 250 255
- Thr Ser Pro Leu Ser Asn His Pro Leu Ala Gly Gly Ser Gly Pro Ser 260 265 270
- Ala Gly Ala Gly Leu Leu Arg Ala Glu Ser Leu Pro Gly Ala Gly Gly 275 280 285
- Ser Leu Thr Arg Thr Pro Leu Met Ser Gln Leu Ile Glu Lys Pro Val
- Ala Pro Ser Val Met Pro Ala Ala Ala Gly Ser Ser Ala Thr Gly 305 310 315
- Gly Ala Ala Pro Val Gly Ala Gly Ala Met Gly Gln Gly Ala Gln Ser 325 330 335
- Gly Gly Ser Thr Arg Pro Gly Leu Val Ala Pro Ala Pro Leu Ala Gln
 340 345
- Glu Arg Glu Glu Asp Asp Glu Asp Asp Trp Asp Glu Glu Asp Asp Trp
 355 360 365

(2) INFORMATION FOR SEQ ID NO:115:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 100 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:
- Met Ala Glu Met Lys Thr Asp Ala Ala Thr Leu Ala Gln Glu Ala Gly

 10 15

WO 99/42076 PCT/US99/03268

134

	Asn	Phe	Glu	Ara	· Tle	Sam	. (7)		_								
							Gly		45					30			
	Glu	. Ser	Thr 35	Ala	Gly	Ser	Leu	Gln 40	Gly	Gln	Trp	Arg	Gly 45	Ala	Ala	Gly	
	Thr	Ala 50	Ala	Gln	Ala	Ala	Val 55	Val	Arg	Phe	Gln	Glu 60	Ala	Ala	Asn	Lys	
	Gln 65	Lys	Gln	Glu	Leu	Asp 70	Glu	Ile	Ser	Thr	Asn 75	Ile	Arg	Gln	Ala	Gly 80	
	Val	Gln	Tyr	Ser	Arg 85	Ala	Asp	Glu	Glu	Gln 90	Gln	Gln	Ala	Leu	Ser 95	Ser	
	Gln	Met	Gly	Phe 100													
(2)	INFO	RMAT	'ION	FOR	SEQ	ID N	0:11	6 :									
((A) (B) (C) (D)	TYP STR TOP	GTH: E: n: ANDE OLOG	396 ucle: DNES: Y: 1:	bas ic a S: s inea	ingle r	irs e									
	xi) : CCGG(
GATCT																	60
AGCAGO	CAAT	` AAG	CAGA	AGC	AGGA	. אריייר	יכת כ	GCCC	AGGC	C GC	GGT	GTGC	: GC1	TTCC	LAGA		120
CGTCC	ATAC	TCG	AGGG	CCG	ACGA	GGZG	CA C	CAGA	rere	G AC	GAAT	ATTC	GTC	AGGC	CGG		180
CTGACC	CGCT	' AAT	'ACGA	AAA	GAAA	CGGZ	יפר ז	ת ת ת ת	AGGC	G CI	GTCC	TCGC	AAA	\TGG@	CTT		240
GCGGGT																	300
CTTGAC										G TC	ACGT	CCAT	TCA	TTCC	CTC		360
(2) IN								1000	n.								396
(i) SI	EQUE: (A) 1 (B) 1 (C) 5 (D) 1	NCE (LENG! TYPE: STRAM	CHARI TH: 8 : am: VDEDN LOGY:	ACTER 30 am ino a NESS:	RIST mino acid sin mear	ICS: acio										
	i) SE																
1	le Se	r Gl	y As	p Le 5	u Ly	s Th	ır Gl	n Il	e As	p Gl	.n Va	il Gl	u Se	er Th		la	

WO 99/42076

Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala 20 25 30											
Ala Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln 35 40 45											
Leu Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr 50 55 60	Ser										
Arg Ala Asp Glu Glu Gln Gln Ala Leu Ser Ser Gln Met Gly 70 75	Phe 80										
(2) INFORMATION FOR SEQ ID NO:118:											
(i) SEQUENCE CHARACTERISTICS:											
(A) LENGTH: 387 base pairs (B) TYPE: nucleic acid											
(C) STRANDEDNESS: single (D) TOPOLOGY: linear											
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:											
GTGGATCCCG ATCCCGTGTT TCGCTATTCT ACGCGAACTC GGCGTTGCCC TATGCGAACA											
TCCCAGTGAC GTTGCCTTCG GTCGAAGCCA TTGCCTGACC GGCTTCGCTG ATCGTCCGCG	60										
	120										
CCAGGTTCTG CAGCGCGTTG TTCAGCTCGG TAGCCGTGGC GTCCCATTTT TGCTGGACAC	180										
CCTGGTACGC CTCCGAACCG CTACCGCCCC AGGCCGCTGC GAGCTTGGTC AGGGACTGCT	240										
TCCCCTCGTC AAGGAGGGAA TGAATGGACG TGACATTTCC CTGGATTGCG CTTGCCGCGG											
CCTCGATACC CGCGAAATTC CACTGCTGCT CTGTCATGTT TTTGCTCCGT TTCTTTTCGT	300										
	360										
ATTAGCGGGT CAGAAGCCCA TTTGCGA	387										
(2) INFORMATION FOR SEQ ID NO:119:											
(i) SEQUENCE CHARACTERISTICS:											
(A) LENGTH: 272 pase pairs											
(B) TYPE: nucleic acid (C) STRANDEDNESS: single											
(D) TOPOLOGY: linear											
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:											
CGGCACGAGG ATCTCGGTTG GCCCAACGGG GCTGGCGAGG GCTCCGTTCC GGGGGCGAGC	60										
TGCGCGCCGG ATGCTTCCTC TGCCCGCAGC CGCGCCTGGA TGGATGGACC AGTTGCTACC	120										
TTCCCGACGT TTCGTTCGGT GTCTGTGCGA TAGCGGTGAC CCCGGCGCGC ACGTCGGGAG	180										
TGTTGGGGGG CAGGCCGGGT CGGTGGTTCG GCCGGGGACG CAGACGGTCT GGACGGAACG	240										

WO 99/42076

GGCGGGGGTT CGCCGATTGG CATCTTTGCC CA

272

- (2) INFORMATION FOR SEQ ID NO:120:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

Asp Pro Val Asp Ala Val Ile Asn Thr Thr Cys Asn Tyr Gly Gln Val

Val Ala Ala Leu 20

- (2) INFORMATION FOR SEQ ID NO:121:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

Ala Val Glu Ser Gly Met Leu Ala Leu Gly Thr Pro Ala Pro Ser

15

- (2) INFORMATION FOR SEQ ID NO:122:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Glu Ala Ala Lys

1 10 15

Glu Gly Arg

- (2) INFORMATION FOR SEQ ID NO:123:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

Tyr Tyr Trp Cys Pro Gly Gln Pro Phe Asp Pro Ala Trp Gly Pro 1

- (2) INFORMATION FOR SEQ ID NO:124:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 14 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

Asp Ile Gly Ser Glu Ser Thr Glu Asp Gln Gln Xaa Ala Val

- (2) INFORMATION FOR SEQ ID NO:125:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

Ala Glu Glu Ser Ile Ser Thr Xaa Glu Xaa Ile Val Pro

- (2) INFORMATION FOR SEQ ID NO:126:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

Asp Pro Glu Pro Ala Pro Pro Val Pro Thr Thr Ala Ala Ser Pro Pro 1 5 10

Ser

- (2) INFORMATION FOR SEQ ID NO:127:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

Ala Pro Lys Thr Tyr Xaa Glu Glu Leu Lys Gly Thr Asp Thr Gly
1 5 10 10 15

- (2) INFORMATION FOR SEQ ID NO:128:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln Leu Thr Ser

1 5 10 15

Leu Leu Asn Ser Leu Ala Asp Pro Asn Val Ser Phe Ala Asn 20 25 30

- (2) INFORMATION FOR SEQ ID NO:129:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Asp Pro Pro Asp Pro His Gln Xaa Asp Met Thr Lys Gly Tyr Tyr Pro 1 5 10 15

Gly Gly Arg Arg Xaa Phe

- (2) INFORMATION FOR SEQ ID NO:130:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 7 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

Asp Pro Gly Tyr Thr Pro Gly

- (2) INFORMATION FOR SEQ ID NO:131:
 - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 10 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear
- (ix) FEATURE:
- (D) OTHER INFORMATION: /note= "The Second Residue Can Be Either a
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

Xaa Xaa Gly Phe Thr Gly Pro Gln Phe Tyr
1 5 10

- (2) INFORMATION FOR SEQ ID NO:132:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 9 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (ix) FEATURE:
- (D) OTHER INFORMATION: /note= "The Third Residue Can Be Either a
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

Xaa Pro Xaa Val Thr Ala Tyr Ala Gly

- (2) INFORMATION FOR SEQ ID NO:133:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 9 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

Xaa Xaa Xaa Glu Lys Pro Phe Leu Arg

- (2) INFORMATION FOR SEQ ID NO:134:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

Xaa Asp Ser Glu Lys Ser Ala Thr Ile Lys Val Thr Asp Ala Ser

- (2) INFORMATION FOR SEQ ID NO:135:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

Ala Gly Asp Thr Xaa Ile Tyr Ile Val Gly Asn Leu Thr Ala Asp

- (2) INFORMATION FOR SEQ ID NO:136:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

Ala Pro Glu Ser Gly Ala Gly Leu Gly Gly Thr Val Gln Ala Gly

15

- (2) INFORMATION FOR SEQ ID NO:137:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

Xaa Tyr Ile Ala Tyr Xaa Thr Thr Ala Gly Ile Val Pro Gly Lys Ile

15

Asn Val His Leu Val

- (2) INFORMATION FOR SEQ ID NO:138:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 882 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

GCAACGCTGT	CGTGGCCTTT	GCGGTGATCG	GTTTCGCCTC	GCTGGCGGTG	GCGGTGGCGG	60
TCACCATCCG	ACCGACCGCG	GCCTCAAAAC	CGGTAGAGGG	ACACCAAAAC	GCCCAGCCAG	120
GGAAGTTCAT	GCCGTTGTTG	CCGACGCAAC	AGCAGGCGCC	GGTCCCGCCG	CCTCCGCCCG	180
ATGATCCCAC	CGCTGGATTC	CAGGGCGGCA	CCATTCCGGC	TGTACAGAAC	GTGGTGCCGC	240
GGCCGGGTAC	CTCACCCGGG	GTGGGTGGGA	CGCCGGCTTC	GCCTGCGCCG	GAAGCGCCGG	300
CCGTGCCCGG	TGTTGTGCCT	GCCCCGGTGC	CAATCCCGGT	CCCGATCATC	ATTCCCCCGT	360
TCCCGGGTTG	GCAGCCTGGA	ATGCCGACCA	TCCCCACCGC	ACCGCCGACG	ACGCCGGTGA	420
CCACGTCGGC	GACGACGCCG	CCGACCACGC	CGCCGACCAC	GCCGGTGACC	ACGCCGCCAA	480
CGACGCCGCC	GACCACGCCG	GTGACCACGC	CGCCAACGAC	GCCGCCGACC	ACGCCGGTGA	540
CCACGCCACC	AACGACCGTC	GCCCCGACGA	CCGTCGCCCC	GACGACGGTC	GCTCCGACCA	600
CCGTCGCCCC	GACCACGGTC	GCTCCAGCCA	CCGCCACGCC	GACGACCGTC	GCTCCGCAGC	660
CGACGCAGCA	GCCCACGCAA	CAACCAACCC	AACAGATGCC	AACCCAGCAG	CAGACCGTGG	720
CCCCGCAGAC	GGTGGCGCCG	GCTCCGCAGC	CGCCGTCCGG	TGGCCGCAAC	GGCAGCGGCG	780
GGGGCGACTT	ATTCGGCGGG	TTCTGATCAC	GGTCGCGGCT	TCACTACGGT	CGGAGGACAT	840
GGCCGGTGAT	GCGGTGACGG	TGGTGCTGCC	CTGTCTCAAC	GA		882

(2) INFORMATION FOR SEQ ID NO:139:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 815 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

CCATCAACCA ACCGCTCGCG	CCGCCCGCGC	CGCCGGATCC	GCCGTCGCCG	CCACGCCCGC	60
CGGTGCCTCC GGTGCCCCCG	TTGCCGCCGT	CGCCGCCGTC	GCCGCCGACC	GGCTGGGTGC	120
CTAGGGCGCT GTTACCGCCC	TGGTTGGCGG	GGACGCCGCC	GGCACCACCG	GTACCGCCGA	180
TGGCGCCGTT GCCGCCGGCG	GCACCGTTGC	CACCGTTGCC	ACCGTTGCCA	CCGTTGCCGA	240
CCAGCCACCC GCCGCGACCA	CCGGCACCGC	CGGCGCCGCC	CGCACCGCCG	GCGTGCCCGT	300
TCGTGCCCGT ACCGCCGGCA	CCGCCGTTGC	CGCCGTCACC	GCCGACGGAA	CTACCGGCGG	360
ACGCGGCCTG CCCGCCGGCG	CCGCCCGCAC	CGCCATTGGC	ACCGCCGTCA	CCGCCGGCTG	420

GGAGTGCCGC	GATTAGGGCA	CTGACCGGCG	CAACCAGCGC	AAGTACTCTC	GGTCACCGAG	480
CACTTCCAGA	CGACACCACA	GCACGGGGTT	GTCGGCGGAC	TGGGTGAAAT	GGCAGCCGAT	540
AGCGGCTAGC	TGTCGGCTGC	GGTCAACCTC	GATCATGATG	TCGAGGTGAC	CGTGACCGCG	600
CCCCCGAAG	GAGGCGCTGA	ACTCGGCGTT	GAGCCGATCG	GCGATCGGTT	GGGGCAGTGC	660
CCAGGCCAAT	ACGGGGATAC	CGGGTGTCNA	AGCCGCCGCG	AGCGCAGCTT	CGGTTGCGCG	720
ACNGTGGTCG	GGGTGGCCTG	TTACGCCGTT	GTCNTCGAAC	ACGAGTAGCA	GGTCTGCTCC	780
GGCGAGGGCA	TCCACCACGC	GTTGCGTCAG	CTCGT			815

(2) INFORMATION FOR SEQ ID NO:140:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1152 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

ACCAGCCGCC GGCTGAGGTC TCAGATCAGA GAGTCTCCGG ACTCACCGGG GCGGTTCAGC 60 CTTCTCCCAG AACAACTGCT GAAGATCCTC GCCCGCGAAA CAGGCGCTGA TTTGACGCTC 120 TATGACCGGT TGAACGACGA GATCATCCGG CAGATTGATA TGGCACCGCT GGGCTAACAG 180 GTGCGCAAGA TGGTGCAGCT GTATGTCTCG GACTCCGTGT CGCGGATCAG CTTTGCCGAC 240 GGCCGGGTGA TCGTGTGGAG CGAGGAGCTC GGCGAGAGCC AGTATCCGAT CGAGACGCTG 300 GACGGCATCA CGCTGTTTGG GCGGCCGACG ATGACAACGC CCTTCATCGT TGAGATGCTC 360 AAGCGTGAGC GCGACATCCA GCTCTTCACG ACCGACGGCC ACTACCAGGG CCGGATCTCA ACACCCGACG TGTCATACGC GCCGCGGCTC CGTCAGCAAG TTCACCGCAC CGACGATCCT 480 GCGTTCTGCC TGTCGTTAAG CAAGCGGATC GTGTCGAGGA AGATCCTGAA TCAGCAGGCC 540 TTGATTCGGG CACACGTC GGGGCAAGAC GTTGCTGAGA GCATCCGCAC GATGAAGCAC 600 TCGCTGGCCT GGGTCGATCG ATCGGGCTCC CTGGCGGAGT TGAACGGGTT CGAGGGAAAT 660 GCCGCAAAGG CATACTTCAC CGCGCTGGGG CATCTCGTCC CGCAGGAGTT CGCATTCCAG 720 GGCCGCTCGA CTCGGCCGCC GTTGGACGCC TTCAACTCGA TGGTCAGCCT CGGCTATTCG 780 CTGCTGTACA AGAACATCAT AGGGGCGATC GAGCGTCACA GCCTGAACGC GTATATCGGT 840

TTCCTACACC	AGGATTCACG	AGGGCACGCA	ACGTCTCGTG	CCGAATTCGG	CACGAGCTCC	900
GCTGAAACCG	CTGGCCGGCT	GCTCAGTGCC	CGTACGTAAT	CCGCTGCGCC	CAGGCCGGCC	960
CGCCGGCCGA	ATACCAGCAG	ATCGGACAGC	GAATTGCCGC	CCAGCCGGTT	GGAGCCGTGC	1020
ATACCGCCGG	CACACTCACC	GGCAGCGAAC	AGGCCTGGCA	CCGTGGCGGC	GCCGGTGTCC	1080
GCGTCTACTT	CGACACCGCC	CATCACGTAG	TGACACGTCG	GCCCGACTTC	CATTGCCTGC	1140
GTTCGGCACG	AG					1152

(2) INFORMATION FOR SEQ ID NO:141:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 655 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

CTCGTGCCGA TTCGGCAGGG	TGTACTTGCC	GGTGGTGTAN	GCCGCATGAG	TGCCGACGAC	60
CAGCAATGCG GCAACAGCAC	GGATCCCGGT	CAACGACGCC	ACCCGGTCCA	CGTGGGCGAT	120
CCGCTCGAGT CCGCCCTGGG	CGGCTCTTTC	CTTGGGCAGG	GTCATCCGAC	GTGTTTCCGC	180
CGTGGTTTGC CGCCATTATG	CCGGCGCGCC	GCGTCGGGCG	GCCGGTATGG	CCGAANGTCG	240
ATCAGCACAC CCGAGATACG	GGTCTGTGCA	AGCTTTTTGA	GCGTCGCGCG	GGGCAGCTTC	300
GCCGGCAATT CTACTAGCGA	GAAGTCTGGC	CCGATACGGA	TCTGACCGAA	GTCGCTGCGG	360
TGCAGCCCAC CCTCATTGGC	GATGGCGCCG	ACGATGGCGC	CTGGACCGAT	CTTGTGCCGC	420
TTGCCGACGG CGACGCGGTA	GGTGGTCAAG	TCCGGTCTAC	GCTTGGGCCT	TTGCGGACGG	480
TCCCGACGCT GGTCGCGGTT	GCGCCGCGAA	AGCGGCGGGT	CGGGTGCCAT	CAGGAATGCC	540
TCACCGCCGC GGCACTGCAC	GGCCAGTGCC	GCGGCGATGT	CAGCCATCGG	GACATCATGC	600
TCGCGTTCAT ACTCCTCGAC	CAGTCGGCGG	AACAGCTCGA	TTCCCGGACC	GCCCA	655

(2) INFORMATION FOR SEQ ID NO:142:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 267 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:
- Asn Ala Val Val Ala Phe Ala Val Ile Gly Phe Ala Ser Leu Ala Val 1 5 10 15
- Ala Val Ala Val Thr Ile Arg Pro Thr Ala Ala Ser Lys Pro Val Glu 20 25 30
- Gly His Gln Asn Ala Gln Pro Gly Lys Phe Met Pro Leu Leu Pro Thr 35 40 45
- Gln Gln Gln Ala Pro Val Pro Pro Pro Pro Pro Asp Asp Pro Thr Ala 50 55 60
- Gly Phe Gln Gly Gly Thr Ile Pro Ala Val Gln Asn Val Val Pro Arg 65 70 75 80
- Pro Gly Thr Ser Pro Gly Val Gly Gly Thr Pro Ala Ser Pro Ala Pro 85 90 95
- Glu Ala Pro Ala Val Pro Gly Val Val Pro Ala Pro Val Pro Ile Pro 100 105 110
- Val Pro Ile Ile Ile Pro Pro Phe Pro Gly Trp Gln Pro Gly Met Pro 115 120 125
- Thr Ile Pro Thr Ala Pro Pro Thr Thr Pro Val Thr Thr Ser Ala Thr 130 135 140
- Thr Pro Pro Thr Thr Pro Pro Thr Thr Pro Val Thr Thr Pro Pro Thr 145
- Thr Pro Pro Thr Thr Pro Val Thr Thr Pro Pro Thr Thr Pro Pro Thr 165 170 175
- Thr Pro Val Thr Thr Pro Pro Thr Thr Val Ala Pro Thr Thr Val Ala 180 185 190
- Pro Thr Thr Val Ala Pro Thr Thr Val Ala Pro Thr Thr Val Ala Pro 195 205
- Ala Thr Ala Thr Pro Thr Thr Val Ala Pro Gln Pro Thr Gln Gln Pro 210 215 220
- Thr Gln Gln Pro Thr Gln Gln Met Pro Thr Gln Gln Gln Thr Val Ala
 225 230 235 240
- Pro Gln Thr Val Ala Pro Ala Pro Gln Pro Pro Ser Gly Gly Arg Asn
 245 250 255
- Gly Ser Gly Gly Gly Asp Leu Phe Gly Gly Phe
- (2) INFORMATION FOR SEQ ID NO:143:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 174 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:
- Ile Asn Gln Pro Leu Ala Pro Pro Ala Pro Pro Asp Pro Pro Ser Pro 1 5 10 15
- Pro Arg Pro Pro Val Pro Pro Val Pro Pro Leu Pro Pro Ser Pro Pro 20 25 30
- Ser Pro Pro Thr Gly Trp Val Pro Arg Ala Leu Leu Pro Pro Trp Leu 35 40
- Ala Gly Thr Pro Pro Ala Pro Pro Val Pro Pro Met Ala Pro Leu Pro 50 55
- Pro Ala Ala Pro Leu Pro Pro Leu Pro Pro Leu Pro Pro Leu Pro Thr 65 70 75 80
- Ser His Pro Pro Arg Pro Pro Ala Pro Pro Ala Pro Pro Ala Pro Pro 85 90 95
- Ala Cys Pro Phe Val Pro Val Pro Pro Ala Pro Pro Leu Pro Pro Ser 100 105 110
- Pro Pro Thr Glu Leu Pro Ala Asp Ala Ala Cys Pro Pro Ala Pro Pro 115 120 125
- Ala Pro Pro Leu Ala Pro Pro Ser Pro Pro Ala Gly Ser Ala Ala Ile 130 135 140
- Arg Ala Leu Thr Gly Ala Thr Ser Ala Ser Thr Leu Gly His Arg Ala 145 150 155 160
- Leu Pro Asp Asp Thr Thr Ala Arg Gly Cys Arg Arg Thr Gly 165 170
- (2) INFORMATION FOR SEQ ID NO:144:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 35 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

Gln Pro Pro Ala Glu Val Ser Asp Gln Arg Val Ser Gly Leu Thr Gly

1 10 15

Ala Val Gln Pro Ser Pro Arg Thr Thr Ala Glu Asp Pro Arg Pro Arg 20 25 30

Asn Arg Arg

- (2) INFORMATION FOR SEQ ID NO:145:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 104 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (xi) SEQUENCE DESCRIPTION: SEO ID NO:145:

Arg Ala Asp Ser Ala Gly Cys Thr Cys Arg Trp Cys Xaa Pro His Glu

1 10 15

Cys Arg Arg Pro Ala Met Arg Gln Gln His Gly Ser Arg Ser Thr Thr 20 25 30

Pro Pro Gly Pro Arg Gly Arg Ser Ala Arg Val Arg Pro Gly Arg Leu 35 40 45

Phe Pro Trp Ala Gly Ser Ser Asp Val Phe Pro Pro Trp Phe Ala Ala 50 55 60

Ile Met Pro Ala Arg Arg Val Gly Arg Pro Val Trp Pro Xaa Val Asp 65 70 75 80

Gln His Thr Arg Asp Thr Gly Leu Cys Lys Leu Phe Glu Arg Arg Ala 85 90 95

Gly Gln Leu Arg Arg Gln Phe Tyr 100

- (2) INFORMATION FOR SEQ ID NO:146:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 53 base pairs
 - (B) TYPE: nucleic acid_
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = "PCR primer"
 - (vi) ORIGINAL SOURCE:

WO 99/42076

		(A) ORGANISM: Mycobacterium tuberculosis	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:146:	
GGAT	CCATA	AT GGGCCATCAT CATCATCATC ACGTGATCGA CATCATCGGG ACC	53
(2)	INFOR	RMATION FOR SEQ ID NO:147:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 42 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: other nucleic acid (A) DESCRIPTION: /desc = "PCR Primer"	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: Mycobacterium tuberculosis	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:147:	
CCTG	GAATT	CA GGCCTCGGTT GCGCCGGCCT CATCTTGAAC GA	42
(2)	INFO	RMATION FOR SEQ ID NO:148:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: other nucleic acid (A) DESCRIPTION: /desc = "PCR Primer"	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: Mycobacterium tuberculosis	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:148:	
GGAT	CCTG	CA GGCTCGAAAC CACCGAGCGG T	31
(2)	INFO	RMATION FOR SEQ ID NO:149:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 31 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: other nucleic acid (A) DESCRIPTION: /desc = "PCR primer"	
	(vi)	ORIGINAL SOURCE:	

(A) ORGANISM: Mycobacterium tuberculosis

(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:	
CTCTGAATTC AGCGCTGGAA ATCGTCGCGA T	31
(2) INFORMATION FOR SEQ ID NO:150:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
<pre>(ii) MOLECULE TYPE: other nucleic acid (A) DESCRIPTION: /desc = "PCR primer"</pre>	
(vi) ORIGINAL SOURCE:(A) ORGANISM: Mycobacterium tuberculosis	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:	
GGATCCAGCG CTGAGATGAA GACCGATGCC GCT	33
(2) INFORMATION FOR SEQ ID NO:151:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 33 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
<pre>(ii) MOLECULE TYPE: other nucleic acid (A) DESCRIPTION: /desc = "PCR primer"</pre>	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Mycobacterium tuberculosis</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:	
GAGAGAATTC TCAGAAGCCC ATTTGCGAGG ACA	33
(2) INFORMATION FOR SEQ ID NO:152:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1993 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: DNA (genomic)	
(vi) ORIGINAL SOURCE:(A) ORGANISM: Mycobacterium tuberculosis	

(A) NAME/KEY: CDS

(B) LOCATION: 152..1273

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

(MI) BIGORNER DESCRIPTION. SEQ ID NO:152:	
TGTTCTTCGA CGGCAGGCTG GTGGAGGAAG GGCCCACCGA ACAGCTGTTC TCCTCGCCGA	60
AGCATGCGGA AACCGCCCGA TACGTCGCCG GACTGTCGGG GGACGTCAAG GACGCCAAGC	120
GCGGAAATTG AAGAGCACAG AAAGGTATGG C GTG AAA ATT CGT TTG CAT ACG Val Lys Ile Arg Leu His Thr 1 5	172
CTG TTG GCC GTG TTG ACC GCT GCG CCG CTG CTG CTA GCA GCG GCG GGC Leu Leu Ala Val Leu Thr Ala Ala Pro Leu Leu Leu Ala Ala Gly 10 15 20	220
TGT GGC TCG AAA CCA CCG AGC GGT TCG CCT GAA ACG GGC GCC GGC GCC Cys Gly Ser Lys Pro Pro Ser Gly Ser Pro Glu Thr Gly Ala Gly Ala 25 30 35	268
GGT ACT GTC GCG ACT ACC CCC GCG TCG TCG CCG GTG ACG TTG GCG GAG Gly Thr Val Ala Thr Thr Pro Ala Ser Ser Pro Val Thr Leu Ala Glu 40 45 50 55	316
ACC GGT AGC ACG CTG CTC TAC CCG CTG TTC AAC CTG TGG GGT CCG GCC Thr Gly Ser Thr Leu Leu Tyr Pro Leu Phe Asn Leu Trp Gly Pro Ala 60 65 70	364
TTT CAC GAG AGG TAT CCG AAC GTC ACG ATC ACC GCT CAG GGC ACC GGT Phe His Glu Arg Tyr Pro Asn Val Thr Ile Thr Ala Gln Gly Thr Gly 75 80 85	412
TCT GGT GCC GGG ATC GCG CAG GCC GCC GGC ACG GTC AAC ATT GGG Ser Gly Ala Gly Ile Ala Gln Ala Ala Gly Thr Val Asn Ile Gly 90 95 100	460
GCC TCC GAC GCC TAT CTG TCG GAA GGT GAT ATG GCC GCG CAC AAG GGG Ala Ser Asp Ala Tyr Leu Ser Glu Gly Asp Met Ala Ala His Lys Gly 105 110 115	508
CTG ATG AAC ATC GCG CTA GCC ATC TCC GCT CAG CAG GTC AAC TAC AAC Leu Met Asn Ile Ala Leu Ala Ile Ser Ala Gln Gln Val Asn Tyr Asn 120 125 130 135	556
CTG CCC GGA GTG AGC GAG CAC CTC AAG CTG AAC GGA AAA GTC CTG GCG Leu Pro Gly Val Ser Glu His Leu Lys Leu Asn Gly Lys Val Leu Ala 140 145 150	604
GCC ATG TAC CAG GGC ACC ATC AAA ACC TGG GAC GAC CCG CAG ATC GCT Ala Met Tyr Gln Gly Thr Ile Lys Thr Trp Asp Asp Pro Gln Ile Ala 155 160 165	652
GCG CTC AAC CCC GGC GTG AAC CTG CCC GGC ACC GCG GTA GTT CCG CTG Ala Leu Asn Pro Gly Val Asn Leu Pro Gly Thr Ala Val Val Pro Leu	700

		170					175					180				
	CGC Arg 185															748
	AAG Lys															796
	GTC Val															844
	GGC Gly															892
	GGC Gly															940
	CAA Gln 265															988
	ATT Ile															1036
_	GCG Ala															1084
	AAC Asn															1132
	GCG Ala															1180
	AAG Lys 345															1228
	GTG Val															1273
															TGCGG	
GCA	TGCT	GGC (CCGT	GCCG	ST G	AAGT	CGGC	C GC	GCTG	GCCC	GGC	CATC	CGG '	TGGT'	TGGGT	1393
GGA	TAGG	TGC	GGTG.	ATCC	CG C'	TGCT	TGCG	C TG	GTCT'	TGGT	GCT	GGTG	GTG	CTGG'	TCATCO	3 1453

AGGG	CGAT	GGG :	rgcgi	ATCA	GG C	rcaa	CGGG:	r TG	CATTI	CTT	CAC	CGCCI	ACC (GAATO	GAAT	C 1513
CAG	GCAA(CAC (CTAC	GCG!	AA AG	CCGT	rgtc <i>i</i>	A CC	BACGO	GTC	GCC	CATC	CGG '	rcggo	GCCT	A 1573
CTA	CGGG	GCG 1	rtgc	CGCT	A TO	CGTC	GGA	G GC	rggco	SACC	TCGC	CAA:	rcg (CCCT	SATCA:	T 1633
CGC	GTG	CCG (STCTO	CTGT	AG GA	AGCGG	GCGCT	r gg:	rgat(CGTG	GAAG	GGC	rgc (CGAAZ	ACGGT	T 1693
GGC	CGAG	GCT (STGG	GAATA	AG TO	CCTGC	GAAT"	r GC	rcgco	CGGA	ATC	CCA	GCG '	rggto	CGTCG	G 1753
TTTC	STGG	GG (GCAAT	rgaco	T TO	CGGG	CCGT	CA:	rcgci	CAT	CACA	ATCG	CTC (CGGT	SATCG	C 1813
TCAG	CAAC	GCT (CCCGI	ATGTO	GC CC	GTG	CTGA	A CTA	ACTTO	GCGC	GGC	FACC	CGG (GCAA	:GGGG	A 1873
GGG	CATG	rtg (GTGT	CCGG	C TO	GTG	TGGG	GG:	rgato	GTC	GTT	CCA	TA 7	rcgco	CACCAC	C 1933
CAC	rca r (GAC (CTGT	rccgo	GC AC	GTG	CCGG1	r GT	rgccc	CCGG	GAGO	GCG	CGA '	rcggo	SAATT	C 1993
(2)	INF	ORMA:	NOI	FOR	SEQ	ID 1	10:15	53:								
		(i) \$	(B)	LEN TYP	NGTH:	: 374 amino	ERIST 4 ami 5 aci 1inea	ino a id		3						
	(:	ii) 1	OLE	TULE	TYPE	E: pi	rotei	in								
	(2	xi) s	SEQUI	ENCE	DES	CRIPT	rion	: SE(O ID	NO:	153:					
Val 1												Thr	Ala	Ala 15	Pro	
1	Lys	Ile	Arg	Leu 5	His	Thr	Leu	Leu	Ala 10	Val	Leu					
1 Leu	Lys Leu	Ile Leu	Arg Ala 20	Leu 5 Ala	His Ala	Thr	Leu Cys	Leu Gly 25	Ala 10 Ser	Val Lys	Leu Pro	Pro	Ser 30	15	Ser	
l Leu Pro	Lys Leu Glu	Ile Leu Thr 35	Arg Ala 20 Gly	Leu 5 Ala Ala	His Ala Gly	Thr Gly Ala	Leu Cys Gly 40	Leu Gly 25 Thr	Ala 10 Ser Val	Val Lys Ala	Leu Pro Thr	Pro Thr 45	Ser 30 Pro	15 Gly	Ser Ser	
l Leu Pro Ser	Lys Leu Glu Pro 50	Ile Leu Thr 35 Val	Arg Ala 20 Gly Thr	Leu 5 Ala Ala Leu	His Ala Gly Ala	Thr Gly Ala Glu 55	Leu Cys Gly 40 Thr	Leu Gly 25 Thr	Ala 10 Ser Val	Val Lys Ala Thr	Leu Pro Thr Leu 60	Pro Thr 45	Ser 30 Pro	15 Gly Ala	Ser Ser Leu	
Leu Pro Ser Phe 65	Lys Leu Glu Pro 50 Asn	Ile Leu Thr 35 Val	Arg Ala 20 Gly Thr	Leu 5 Ala Ala Leu Gly	His Ala Gly Ala Pro 70	Thr Gly Ala Glu 55	Leu Cys Gly 40 Thr	Leu Gly 25 Thr Gly	Ala 10 Ser Val Ser	Val Lys Ala Thr	Leu Pro Thr Leu 60	Pro Thr 45 Leu Pro	Ser 30 Pro Tyr Asn	15 Gly Ala Pro	Ser Ser Leu Thr 80	
Leu Pro Ser Phe 65	Lys Leu Glu Pro 50 Asn	Ile Leu Thr 35 Val Leu Ala	Arg Ala 20 Gly Thr Trp	Leu 5 Ala Ala Leu Gly Gly 85	His Ala Gly Ala Pro 70 Thr	Thr Gly Ala Glu 55 Ala Gly	Leu Cys Gly 40 Thr Phe	Leu Gly 25 Thr Gly His	Ala 10 Ser Val Ser Glu Ala 90	Val Lys Ala Thr Arg 75	Leu Pro Thr Leu 60 Tyr	Pro Thr 45 Leu Pro	Ser 30 Pro Tyr Asn	Gly Ala Pro Val	Ser Leu Thr 80	

Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His Leu Lys

	130					135					140				
Leu 145	Asn	Gly	Lys	Val	Leu 150	Ala	Ala	Met	Tyr	Gln 155	Gly	Thr	Ile	Lys	Thr 160
Trp	Asp	Asp	Pro	Gln 165	Ile	Ala	Ala	Leu	Asn 170	Pro	Gly	Val	Asn	Leu 175	Pro
Gly	Thr	Ala	Val 180	Val	Pro	Leu	His	Arg 185	Ser	Asp	Gly	Ser	Gly 190	Asp	Thi
Phe	Leu	Phe 195	Thr	Gln	Tyr	Leu	Ser 200	Lys	Gln	Asp	Pro	Glu 205	Gly	Trp	Gly
Lys	Ser 210	Pro	Gly	Phe	Gly	Thr 215	Thr	Val	Asp	Phe	Pro 220	Ala	Val	Pro	Gly
Ala 225	Leu	Gly	Glu	Asn	Gly 230	Asn	Gly	Gly	Met	Val 235	Thr	Gly	Cys	Ala	Glu 240
Thr	Pro	Gly	Cys	Val 245	Ala	Tyr	Ile	Gly	Ile 250	Ser	Phe	Leu	Asp	Gln 255	Ala
Ser	Gln	Arg	Gly 260	Leu	Gly	Glu	Ala	Gln 265	Leu	Gly	Asn	Ser	Ser 270	Gly	Asr
Phe	Leu	Leu 275	Pro	Asp	Ala	Gln	Ser 280	Ile	Gln	Ala	Ala	Ala 285	Ala	Gly	Phe
Ala	Ser 290	Lys	Thr	Pro	Ala	Asn 295	Gln	Ala	Ile	Ser	Met 300	Ile	Asp	Gly	Pro
Ala 305	Pro	Asp	Gly	Tyr	Pro 310	Ile	Ile	Asn	Tyr	Glu 315	Tyr	Ala	Ile	Val	Asi 320
Asn	Arg	Gln	Lys	Asp 325	Ala	Ala	Thr	Ala	Gln 330	Thr	Leu	Gln	Ala	Phe 335	Let
His	Trp	Ala	Ile 340	Thr	Asp	Gly	Asn	Lys 345	Ala	Ser	Phe	Leu	Asp 350	Gln	Va:
His	Phe	Gln 355	Pro	Leu	Pro	Pro	Ala 360	Val	Val	Lys	Leu	Ser 365	Asp	Ala	Le
Ile	Ala	Thr	Ile	Ser	Ser										

- (2) INFORMATION FOR SEQ ID NO:154:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1993 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

TGTTCTTCGA	CGGCAGGCTG	GTGGAGGAAG	GGCCCACCGA	ACAGCTGTTC	TCCTCGCCGA	60
AGCATGCGGA	AACCGCCCGA	TACGTCGCCG	GACTGTCGGG	GGACGTCAAG	GACGCCAAGC	120
GCGGAAATTG	AAGAGCACAG	AAAGGTATGG	CGTGAAAATT	CGTTTGCATA	CGCTGTTGGC	180
CGTGTTGACC	GCTGCGCCGC	TGCTGCTAGC	AGCGGCGGGC	TGTGGCTCGA	AACCACCGAG	240
CGGTTCGCCT	GAAACGGGCG	CCGGCGCCGG	TACTGTCGCG	ACTACCCCCG	CGTCGTCGCC	300
GGTGACGTTG	GCGGAGACCG	GTAGCACGCT	GCTCTACCCG	CTGTTCAACC	TGTGGGGTCC	360
GGCCTTTCAC	GAGAGGTATC	CGAACGTCAC	GATCACCGCT	CAGGGCACCG	GTTCTGGTGC	420
CGGGATCGCG	CAGGCCGCCG	CCGGGACGGT	CAACATTGGG	GCCTCCGACG	CCTATCTGTC	480
GGAAGGTGAT	ATGGCCGCGC	ACAAGGGGCT	GATGAACATC	GCGCTAGCCA	TCTCCGCTCA	540
GCAGGTCAAC	TACAACCTGC	CCGGAGTGAG	CGAGCACCTC	AAGCTGAACG	GAAAAGTCCT	600
GGCGGCCATG	TACCAGGGCA	CCATCAAAAC	CTGGGACGAC	CCGCAGATCG	CTGCGCTCAA	660
CCCCGGCGTG	AACCTGCCCG	GCACCGCGGT	AGTTCCGCTG	CACCGCTCCG	ACGGGTCCGG	720
TGACACCTTC	TTGTTCACCC	AGTACCTGTC	CAAGCAAGAT	CCCGAGGGCT	GGGGCAAGTC	780
GCCCGGCTTC	GGCACCACCG	TCGACTTCCC	GGCGGTGCCG	GGTGCGCTGG	GTGAGAACGG	840
CAACGGCGGC	ATGGTGACCG	GTTGCGCCGA	GACACCGGGC	TGCGTGGCCT	ATATCGGCAT	900
CAGCTTCCTC	GACCAGGCCA	GTCAACGGGG	ACTCGGCGAG	GCCCAACTAG	GCAATAGCTC	960
TGGCAATTTC	TTGTTGCCCG	ACGCGCAAAG	CATTCAGGCC	GCGGCGGCTG	GCTTCGCATC	1020
GAAAACCCCG	GCGAACCAGG	CGATTTCGAT	GATCGACGGG	CCCGCCCCGG	ACGGCTACCC	1080
GATCATCAAC	TACGAGTACG	CCATCGTCAA	CAACCGGCAA	AAGGACGCCG	CCACCGCGCA	1140
GACCTTGCAG	GCATTTCTGC	ACTGGGCGAT	CACCGACGGC	AACAAGGCCT	CGTTCCTCGA	1200
CCAGGTTCAT	TTCCAGCCGC	TGCCGCCCGC	GGTGGTGAAG	TTGTCTGACG	CGTTGATCGC	1260
GACGATTTCC	AGCTAGCCTC	GTTGACCACC	ACGCGACAGC	AACCTCCGTC	GGGCCATCGG	1320
GCTGCTTTGC	GGAGCATGCT	GGCCCGTGCC	GGTGAAGTCG	GCCGCGCTGG	CCCGGCCATC	1380
CGGTGGTTGG	GTGGGATAGG	TGCGGTGATC	CCGCTGCTTG	CGCTGGTCTT	GGTGCTGGTG	1440
GTGCTGGTCA	TCGAGGCGAT	GGGTGCGATC	AGGCTCAACG	GGTTGCATTT	CTTCACCGCC	1500
ACCGAATGGA	ATCCAGGCAA	CACCTACGGC	GAAACCGTTG	TCACCGACGC	GTCGCCCATC	1560
CGGTCGGCGC	CTACTACGGG	GCGTTGCCGC	TGATCGTCGG	GACGCTGGCG	ACCTCGGCAA	1620

TCGCCCTG.	AT C	ATCG	CGGTC	CCC	GTCI	CTG	TAGG	SAGCO	GC (CTG	TGAT	C GI	GGA	ACGGC	2	1680
TGCCGAAA	CG G	rtggo	CCGAG	GCT	GTGG	GAA	TAGT	CCT	GA I	ATTGO	TCGC	C GC	BAATO	CCC	À.	1740
GCGTGGTC	GT C	GGTTT	GTGG	GGG	GCA	ATGA	CGTT	CGGG	CC C	STTCA	TCGC	T CA	ATCAC	ATC	5	1800
CTCCGGTG.	AT C	GCTC	CAAC	GCI	ccc	SATG	TGCC	CGGTC	CT (SAACT	ACTI	G CC	GCGGC	GAC	:	1860
CGGGCAAC	GG GO	GAGGO	CATO	TTC	GTGT	CCG	GTCT	GGT	TT C	GCGG	TGAT	G G	CGT	CCCF	1	1920
TTATCGCC	AC C	ACCAC	TCAT	GAC	CTGI	TCC	GGCA	GGT	CC C	GTGT	TGCC	ec co	GGAG	GGC	;	1980
CGATCGGG.	AA T	rc														1993
(2) INFO	RMAT	ION I	OR S	SEQ I	D NC):155	5 :									
	(B)	LEN TYI STI TOI	NGTH: PE: a RANDE POLOG	374 mino EDNES	ami aci SS: inea	ino a ld ar	cids		155							
Met 1	Lys	ITE	Arg	Leu 5	His	Thr	Leu	Leu	Ala 10	Val	Leu	Thr	Ala	Ala 15	Pro	
Leu	Leu	Leu	Ala 20	Ala	Ala	Gly	Cys	Gly 25	Ser	Lys	Pro	Pro	Ser 30	Gly	Ser	
Pro	Glu	Thr 35	Gly	Ala	Gly	Ala	Gly 40	Thr	Val	Ala	Thr	Thr 45	Pro	Ala	Ser	
Ser	Pro 50	Val	Thr	Leu	Ala	Glu 55	Thr	Gly	Ser	Thr	Leu 60	Leu	Tyr	Pro	Leu	
Phe 65	Asn	Leu	Trp	Gly	Pro 70	Ala	Phe	His	Glu	Arg 75	Tyr	Pro	Asn	Val	Thr 80	
Ile	Thr	Ala	Gln	Gly 85	Thr	Gly	Ser	Gly	Ala 90	Gly	Ile	Ala	Gln	Ala 95	Ala	
Ala	Gly	Thr	Val 100	Asn	Ile	Gly	Ala	Ser 105	Asp	Ala	Tyr	Leu	Ser 110	Glu	Gly	
Asp	Met	Ala 115	Ala	His	Lys	Gly	Leu 120	Met	Asn	Ile	Ala	Leu 125	Ala	Ile	Ser	
Ala	Gln 130	Gln	Val	Asn	Tyr	Asn 135	Leu	Pro	Gly	Val	Ser 140	Glu	His	Leu	Lys	
Leu 145	Asn	Gly	Lys	Val	Leu 150	Ala	Ala	Met	Tyr	Gln 155	Gly	Thr	Ile	Lys	Thr 160	

WO 99/42076 PCT/US99/03268

155

Trp	Asp	Asp	Pro	Gln	Ile	Ala	Ala	Leu	Asn	Pro	Gly	Val	Asn	Leu	Pro
				165					170					175	

- Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr 180 185 190
- Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly 195 200 205
- Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly 210 215 220
- Ala Leu Gly Glu Asn Gly Asn Gly Gly Met Val Thr Gly Cys Ala Glu 225 230 235 240
- Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala 245 250 255
- Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser Gly Asn 260 265 270
- Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe 275 280 285
- Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro 290 295 300
- Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile Val Asn 305 310 315 320
- Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu 325 330 335
- His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp Gln Val
- His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp Ala Leu 355 360 365
- Ile Ala Thr Ile Ser Ser

(2) INFORMATION FOR SEQ ID NO:156:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1777 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

GGTCTTGACC ACCACCTGGG TGTCGAAGTC GGTGCCCGGA TTGAAGTCCA GGTACTCGTG

GTAGCGAAAC	GGCAACGCGG	CCGCGTTGGG	CACCTTGTTC	AGCGCTGATT	TGCACAACAC	180
CTCGTGGAAG	GTGATGCCGT	CGAATTGTGG	CGCGCGAACG	CTGCGGACCA	GGCCGATCCG	240
CTGCAACCCG	GCAGCGCCCG	TCGTCAACGG	GCATCCCGTT	CACCGCGACG	GCTTGCCGGG	300
CCCAACGCAT	ACCATTATTC	GAACAACCGT	TCTATACTTT	GTCAACGCTG	GCCGCTACCG	360
AGCGCCGCAC	AGGATGTGAT	ATGCCATCTC	TGCCCGCACA	GACAGGAGCC	AGGCCTTATG	420
ACAGCATTCG	GCGTCGAGCC	CTACGGGCAG	CCGAAGTACC	TAGAAATCGC	CGGGAAGCGC	480
ATGGCGTATA	TCGACGAAGG	CAAGGGTGAC	GCCATCGTCT	TTCAGCACGG	CAACCCCACG	540
TCGTCTTACT	TGTGGCGCAA	CATCATGCCG	CACTTGGAAG	GGCTGGGCCG	GCTGGTGGCC	600
TGCGATCTGA	TCGGGATGGG	CGCGTCGGAC	AAGCTCAGCC	CATCGGGACC	CGACCGCTAT	660
AGCTATGGCG	AGCAACGAGA	CTTTTTGTTC	GCGCTCTGGG	ATGCGCTCGA	CCTCGGCGAC	720
CACGTGGTAC	TGGTGCTGCA	CGACTGGGGC	TCGGCGCTCG	GCTTCGACTG	GGCTAACCAG	780
CATCGCGACC	GAGTGCAGGG	GATCGCGTTC	ATGGAAGCGA	TCGTCACCCC	GATGACGTGG	840
GCGGACTGGC	CGCCGGCCGT	GCGGGGTGTG	TTCCAGGGTT	TCCGATCGCC	TCAAGGCGAG	900
CCAATGGCGT	TGGAGCACAA	CATCTTTGTC	GAACGGGTGC	TGCCCGGGGC	GATCCTGCGA	960
CAGCTCAGCG	ACGAGGAAAT	GAACCACTAT	CGGCGGCCAT	TCGTGAACGG	CGGCGAGGAC	1020
CGTCGCCCCA	CGTTGTCGTG	GCCACGAAAC	CTTCCAATCG	ACGGTGAGCC	CGCCGAGGTC	1080
GTCGCGTTGG	TCAACGAGTA	CCGGAGCTGG	CTCGAGGAAA	CCGACATGCC	GAAACTGTTC	1140
ATCAACGCCG	AGCCCGGCGC	GATCATCACC	GGCCGCATCC	GTGACTATGT	CAGGAGCTGG	1200
CCCAACCAGA	CCGAAATCAC	AGTGCCCGGC	GTGCATTTCG	TTCAGGAGGA	CAGCGATGGC	1260
GTCGTATCGT	GGGCGGCGC	TCGGCAGCAT	CGGCGACCTG	GGAGCGCTCT	CATTTCACGA	1320
GACCAAGAAT	GTGATTTCCG	GCGAAGGCGG	CGCCCTGCTT	GTCAACTCAT	AAGACTTCCT	1380
GCTCCGGGCA	GAGATTCTCA	GGGAAAAGGG	CACCAATCGC	AGCCGCTTCC	TTCGCAACGA	1440
GGTCGACAAA	TATACGTGGC	AGGACAAAGG	TCTTCCTATT	TGCCCAGCGA	ATTAGTCGCT	1500
GCCTTTCTAT	GGGCTCAGTT	CGAGGAAGCC	GAGCGGATCA	CGCGTATCCG	ATTGGACCTA	1560
TGGAACCGGT	ATCATGAAAG	CTTCGAATCA	TTGGAACAGC	GGGGGCTCCT	GCGCCGTCCG	1620
ATCATCCCAC	AGGGCTGCTC	TCACAACGCC	CACATGTACT	ACGTGTTACT	AGCGCCCAGC	1680
GCCGATCGGG	AGGAGGTGCT	GGCGCGTCTG	ACGAGCGAAG	GTATAGGCGC	GGTCTTTCAT	1740

WO 99/42076

TACGTGCCGC TTCACGATTC GCCGGCCGGG CGTCGCT	1777
(2) INFORMATION FOR SEQ ID NO:157:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 324 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:	
GAGATTGAAT CGTACCGGTC TCCTTAGCGG CTCCGTCCCG TGAATGCCCA TATCACGCAC	60
GGCCATGTTC TGGCTGTCGA CCTTCGCCCC ATGCCCGGAC GTTGGTAAAC CCAGGGTTTG	120
ATCAGTAATT CCGGGGGACG GTTGCGGGAA GGCGGCCAGG ATGTGCGTGA GCCGCGGCGC	180
CGCCGTCGCC CAGGCGACCG CTGGATGCTC AGCCCCGGTG CGGCGACGTA GCCAGCGTTT	240
GGCGCGTGTC GTCCACAGTG GTACTCCGGT GACGACGCGG CGCGGTGCCT GGGTGAAGAC	300
CGTGACCGAC GCCGCCGATT CAGA	324
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1338 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:	
GCGGTACCGC CGCGTTGCGC TGGCACGGGA CCTGTACGAC CTGAACCACT TCGCCTCGCG	60
AACGATTGAC GAACCGCTCG TGCGGCGGCT GTGGGTGCTC AAGGTGTGGG GTGATGTCGT	120
CGATGACCGG CGCGGCACCC GGCCACTACG CGTCGAAGAC GTCCTCGCCG CCCGCAGCGA	180
GCACGACTTC CAGCCCGACT CGATCGGCGT GCTGACCCGT CCTGTCGCTA TGGCTGCCTG	240
GGAAGCTCGC GTTCGGAAGC GATTTGCGTT CCTCACTGAC CTCGACGCCG ACGAGCAGCG	300
GTGGGCCGCC TGCGACGAAC GGCACCGCCG CGAAGTGGAG AACGCGCTGG CGGTGCTGCG	360
GTCCTGATCA ACCTGCCGGC GATCGTGCCG TTCCGCTGGC ACGGTTGCGG CTGGACGCGG	420
CTGAATCGAC TAGATGAGAG CAGTTGGGCA €GAATCCGGC TGTGGTGGTG AGCAAGACAC	480
GAGTACTGTC ATCACTATTG GATGCACTGG ATGACCGGCC TGATTCAGCA GGACCAATGG	540
AACTGCCCGG GGCAAAACGT CTCGGAGATG ATCGGCGTCC CCTCGGAACC CTGCGGTGCT	600
GGCGTCATTC GGACATCGGT CCGGCTCGCG GGATCGTGGT GACGCCAGCG CTGAAGGAGT	660

GGAGCGCGGC	GGTGCACGCG	CTGCTGGACG	GCCGGCAGAC	GGTGCTGCTG	CGTAAGGGCG	720
GGATCGGCGA	GAAGCGCTTC	GAGGTGGCGG	CCCACGAGTT	CTTGTTGTTC	CCGACGGTCG	780
CGCACAGCCA	CGCCGAGCGG	GTTCGCCCCG	AGCACCGCGA	CCTGCTGGGC	CCGGCGGCCG	840
CCGACAGCAC	CGACGAGTGT	GTGCTACTGC	GGGCCGCAGC	GAAAGTTGTT	GCCGCACTGC	900
CGGTTAACCG	GCCAGAGGGT	CTGGACGCCA	TCGAGGATCT	GCACATCTGG	ACCGCCGAGT	960
CGGTGCGCGC	CGACCGGCTC	GACTTTCGGC	CCAAGCACAA	ACTGGCCGTC	TTGGTGGTCT	1020
CGGCGATCCC	GCTGGCCGAG	CCGGTCCGGC	TGGCGCGTAG	GCCCGAGTAC	GGCGGTTGCA	1080
CCAGCTGGGT	GCAGCTGCCG	GTGACGCCGA	CGTTGGCGGC	GCCGGTGCAC	GACGAGGCCG	1140
CGCTGGCCGA	GGTCGCCGCC	CGGGTCCGCG	AGGCCGTGGG	TTGACTGGGC	GGCATCGCTT	1200
GGGTCTGAGC	TGTACGCCCA	GTCGGCGCTG	CGAGTGATCT	GCTGTCGGTT	CGGTCCCTGC	1260
TGGCGTCAAT	TGACGGCGCG	GGCAACAGCA	GCATTGGCGG	CGCCATCCTC	CGCGCGGCCG	1320
GCGCCCACCG	CTACAACC					1338

(2) INFORMATION FOR SEQ ID NO:159:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 321 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:

CCGGCGCAC CGGCGCACC GGCGG	TACCG GCGGCAACGG	CGCTGACGCC	GCTGCTGTGG	60
TGGGCTTCGG CGCGAACGGC GACCC	TGGCT TCGCTGGCGG	CAAAGGCGGT	AACGGCGGAA	120
TAGGTGGGC CGCGGTGACA GGCGG	GGTCG CCGGCGACGG	CGGCACCGGC	GGCAAAGGTG	180
GCACCGGCGG TGCCGGCGGC GCCGG	CAACG ACGCCGGCAG	CACCGGCAAT	CCCGGCGGTA	240
AGGGCGGCGA CGGCGGGATC GGCGG	TGCCG GCGGGGCCGG	CGGCGCGGCC	GGCACCGGCA	300
ACGGCGGCCA TGCCGGCAAC C				321

(2) INFORMATION FOR SEQ ID NO:160:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 492 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(λ	1)	SEQU	JENCE	E DES	SCRI	PTION	1: SI	EQ II	ON C	16 0	:					
GAAGAC	CCG	G CC	CCGC	CATA	A TCC	SATCO	GCT	CGC	CGACI	TAC 1	TTC	GCCG/	AA C	GTGC#	ACGC	3
GCGGCG	TCG	G GC	TGAT	CATO	CAC	CGGTC	GCT	ACG	CGCCC	CAA (CCGC	ACCG	GA TO	GCT	CTG	2
CGTTCG	CCT	C CG	SAACI	CGT	C ACT	rtcgo	GCGC	AAGO	CCCG	ACG (GCACC	GCC	GA AT	CAC	LAGGO	3
CGGTCC	ACG	A TI	CGGC	STGC	A AA	SATCO	CTGC	TGC	TAAL	CCT (GCACC	CCG	GA CO	GCTAC	CGCCI	ŗ
ACCACC	CAC	т то	cggr	CAG	C GCC	CTCGC	CCGA	TCA	AGGCC	GCC (GATC	ACCC	CG TT	rtcgi	rccgo	2
GAGCACTATC GGCTCGCGGG GTCGAAGCGA CCATCGCGGA TTTCGCCCGC TGCGCGCAGT																
TGGCCC	GCG.	A TO	CCGC	CTAC	C GAC	CGGCC	STCG	AAA	CATO	GG (CAGCO	AAGO	G TA	ATCTO	CTC	À.
ATCAGT	TCC	T GO	GCGCC	CGCGC	ACC	CAACA	AAGC	GCA	CCGAC	CTC (STGGG	GCG	GC A	CACCO	GCCI	A.
ACCGTC	GCC	G GI	?													
(2) IN	FOR	MATI	ON E	FOR S	SEQ I	D NO	0:161	L:								
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 536 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: (D) TOPOLOGY: linear 																
(x	i)	SEQU	JENCE	E DES	CRII	PTICE	1: SI	EQ II	NO:	161	:					
	he .		JENCE Gln									Glu	Leu	Trp	Arg 15	Ala
P: 1	he .	Ala		His	Leu 5	Val	Glu	Gly	Asp	Ala 10	Val			_	15	
P: 1 A	he .	Ala Ala	Gln	His Asp 20	Leu 5 Gln	Val Ala	Glu Asp	Gly Pro	Asp Leu 25	Ala 10 Gln	Val Pro	Gly	Ser	Ala 30	15 Arg	Arg
P 1 A G	he sn ln	Ala Ala Arg	Gln Ala Ala	His Asp 20 Ser	Leu 5 Gln Arg	Val Ala Ser	Glu Asp Pro	Gly Pro Arg 40	Asp Leu 25 Arg	Ala 10 Gln Leu	Val Pro Ala	Gly Gly	Ser Pro 45	Ala 30 Asn	15 Arg Ala	Arg Tyr
P. A G H	he sn ln	Ala Ala Arg Tyr 50	Gln Ala Ala 35	His Asp 20 Ser Asn	Leu 5 Gln Arg Asn	Val Ala Ser Arg	Glu Asp Pro Ser 55	Gly Pro Arg 40	Asp Leu 25 Arg Leu	Ala 10 Gln Leu Cys	Val Pro Ala Gln	Gly Gly Arg	Ser Pro 45 Trp	Ala 30 Asn Pro	15 Arg Ala Leu	Arg Tyr Pro
P. A G H	he sn ln is	Ala Ala Arg Tyr 50 Ala	Gln Ala Ala 35 Ser	Asp 20 Ser Asn	Leu 5 Gln Arg Asn	Val Ala Ser Arg Val 70	Glu Asp Pro Ser 55	Gly Pro Arg 40 Ile Cys	Asp Leu 25 Arg Leu	Ala 10 Gln Leu Cys	Val Pro Ala Gln Cys 75	Gly Gly Arg 60 Pro	Ser Pro 45 Trp	Ala 30 Asn Pro	Arg Ala Leu Gln	Arg Tyr Pro Glu 80
P. A. G. H. S. 6	he sn is is er s	Ala Ala Arg Tyr 50 Ala	Gln Ala Ala 35 Ser	Asp 20 Ser Asn Gln	Leu 5 Gln Arg Asn Asp	Val Ala Ser Arg Val 70 Ala	Glu Asp Pro Ser 55 Ile	Gly Pro Arg 40 Ile Cys	Asp Leu 25 Arg Leu His	Ala 10 Gln Leu Cys Leu Glu 90	Val Pro Ala Gln Cys 75 Pro	Gly Gly Arg 60 Pro	Ser Pro 45 Trp His	Ala 30 Asn Pro Arg	Arg Ala Leu Gln Pro 95	Arg Tyr Pro Glu 80 Lys
P 1 A G H S 6	he sn ln lis er s	Ala Arg Tyr 50 Ala Gly Leu	Gln Ala Ala 35 Ser Ala Leu	Asp 20 Ser Asn Gln Met	Leu 5 Gln Arg Asn Asp Thr 85	Val Ala Ser Arg Val 70 Ala Gly	Glu Asp Pro Ser 55 Ile Phe Lys	Gly Pro Arg 40 Ile Cys Gly Arg	Asp Leu 25 Arg Leu His Val	Ala 10 Gln Leu Cys Leu Glu 90 Ala	Val Pro Ala Gln Cys 75 Pro	Gly Gly Arg 60 Pro Tyr	Ser Pro 45 Trp His Gly Asp	Ala 30 Asn Pro Arg Gln Glu 110	Arg Ala Leu Gln Pro 95 Gly	Arg Tyr Pro Glu 80 Lys

	130)				135	i				140				
Cys 145	Asp	Leu	ı Ile	: Gly	Met 150	Gly	Ala	Ser	Asp	Lys 155	Leu	Ser	Pro	Ser	Gly 160
Pro	Asp	Arg	Tyr	Ser 165	Tyr	Gly	Glu	Gln	Arg 170		Phe	Leu	Phe	Ala 175	Leu
Trp	Asp	Ala	Leu 180	Asp	Leu	Gly	Asp	His 185	Val	Val	Leu	Val	Leu 190	His	Asp
Trp	Gly	Ser 195	Ala	Leu	Gly	Phe	Asp 200	Trp	Ala	Asn	Gln	His 205	Arg	Asp	Arg
Val	Gln 210	Gly	Ile	Ala	Phe	Met 215	Glu	Ala	Ile	Val	Thr 220	Pro	Met	Thr	Trọ
Ala 225	Asp	Trp	Pro	Pro	Ala 230	Val	Arg	Gly	Val	Phe 235	Gln	Gly	Phe	Arg	Ser 240
Pro	Gln	Gly	Glu	Pro 245	Met	Ala	Leu	Glu	His 250	Asn	Ile	Phe	Val	Glu 255	Arg
Val	Leu	Pro	Gly 260	Ala	Ile	Leu	Arg	Gln 265	Leu	Ser	Asp	Glu	Glu 270	Met	Asn
His	Tyr	Arg 275	Arg	Pro	Phe	Val	Asn 280	Gly	Gly	Glu	Asp	Arg 285	Arg	Pro	Thr
Leu	Ser 290	Trp	Pro	Arg	Asn	Leu 295	Pro	Ile	As p	Gly	Glu 300	Pro	Ala	Glu	Val
305			Val		310					315					320
			Phe	325					330					335	
			Tyr 340					345					350		
		355	His				360					365			
Ala	Gly 370	Ala	Arg	Gln	His	Arg 375	Arg	Pro	Gly	Ser	Ala 380	Leu	Ile	Ser	Arg
Asp 385	Gln	Glu	Cys	Asp	Phe 390	Arg	Arg	Arg	Arg	Arg 395	Pro	Ala	Cys	Gln	Leu 40 0
Ile	Arg	Leu	Pro	Ala 405	Pro	Gly	Arg	Asp	Ser 410	Gln	Gly	Lys	Gly	His 415	Gln
Ser	Gln	Pro	Leu 420	Pro	Ser	Gln	Arg	Gly 425	Arg	Gln	Ile	Tyr	Val 430	Ala	Gly

- Gln Arg Ser Ser Tyr Leu Pro Ser Glu Leu Val Ala Ala Phe Leu Trp
 435
 440
 445
- Ala Gln Phe Glu Glu Ala Glu Arg Ile Thr Arg Ile Arg Leu Asp Leu
 450
 450
 460
- Trp Asn Arg Tyr His Glu Ser Phe Glu Ser Leu Glu Gln Arg Gly Leu 465 470 475 480
- Leu Arg Arg Pro Ile Ile Pro Gln Gly Cys Ser His Asn Ala His Met
- Tyr Tyr Val Leu Leu Ala Pro Ser Ala Asp Arg Glu Glu Val Leu Ala 500 505 510
- Arg Leu Thr Ser Glu Gly Ile Gly Ala Val Phe His Tyr Val Pro Leu 515 520 525
- His Asp Ser Pro Ala Gly Arg Arg

(2) INFORMATION FOR SEQ ID NO:162:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 284 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:
- Asn Glu Ser Ala Pro Arg Ser Pro Met Leu Pro Ser Ala Arg Pro Arg
- Tyr Asp Ala Ile Ala Val Leu Leu Asn Glu Met His Ala Gly His Cys
 20 25 30
- Asp Phe Gly Leu Val Gly Pro Ala Pro Asp Ile Val Thr Asp Ala Ala 35 40 45
- Gly Asp Asp Arg Ala Gly Leu Gly Val Asp Glu Gln Phe Arg His Val 50 55 60
- Gly Phe Leu Glu Pro Ala Pro Val Leu Val Asp Gln Arg Asp Asp Leu

 70 75
- Gly Gly Leu Thr Val Asp Trp Lys Val Ser Trp Pro Arg Gln Arg Gly 85 90 95
- Ala Thr Val Leu Ala Ala Val His Glu Trp Pro Pro Ile Val Val His
- Phe Leu Val Ala Glu Leu Ser Gln Asp Arg Pro Gly Gln His Pro Phe

	Asp	Lys 130	Asp	Val	Val	Leu	Gln 135	Arg	His	Trp	Leu	Ala 140	Leu	Arg	Arg	Ser	
	Glu 145	Thr	Leu	Glu	His	Thr 150	Pro	His	Gly	Arg	Arg 155	Pro	Val	Arg	Pro	Arg 160	
	His	Arg	Gly	Asp	Asp 165	Arg	Phe	His	Glu	Arg 170	Asp	Pro	Leu	His	Ser 175	Val	
	Ala	Met	Leu	Val 180	Ser	Pro	Val	Glu	Ala 185	Glu	Arg	Arg	Ala	Pro 190	Val	Val	
	Gln	His	Gln 195	Tyr	His	Val	Val	Ala 200	Glu	Val	Glu	Arg	Ile 205	Pro	Glu	Arg	
	Glu	Gln 210	Lys	Val	Ser	Leu	Leu 215	Ala	Ile	Ala	Ile	Ala 220	Val	Gly	Ser	Arg	
	Trp 225	Ala	Glu	Leu	Val	Arg 230	Arg	Ala	His	Pro	As p 235	Gln	Ile	Ala	Gly	His 240	
	Gln	Pro	Ala	Gln	Pro 245	Phe	Gln	Val	Arg	His 250	Asp	Val	Ala	Pro	Gln 255	Val	
	Arg	Arg	Arg	Gly 260	Val	Ala	Val	Leu	Lys 265	Asp	Asp	Gly	Val	Thr 270	Leu	Ala	
	Phe	Val	Asp 275	Ile	Arg	His	Ala	Leu 280	Pro	Gly	Asp	Phe					
(2)	INFOR	MATI	ON F	OR S	EQ I	D NO	:163	:									
	(i)	(A) (B) (C)	ENCE LEN TYP STR TOP	GTH: E: n ANDE	264 ucle DNES	bas ic a S: s	e pa cid ingl	irs									
	(xi)	SEQU	ENCE	DES	CRIP	TION	: SE	Q ID	NO:	163:							
ATGA	ACATG	T CG	TCGG	TGGT	GGG	TCGC	AAG	GCCT	TTGC	GC G	ATTC	GCCG	G CT	ACTO	CTCC	•	60
GCCA	TGCAC	G CG	ATCG	CCGG	TTT	CICC	GAT	GCGT	TGCG	CC A	AGAG	CTGC	G GG	GTAG	CGGA		120
ATCG	CCGTC	r cg	GTGA	TCCA	ccc	GGCG	CTG	ACCC	AGAC	AC C	GCTG	TTGG	c ca	ACGT	CGAC		180
CCCG	CCGAC	A TG	CCGC	CGCC	GTT	TCGC	AGC	CTCA	.cgcc	CA T	TCCC	GTTC	A CT	GGGT	CGCG		240
GCAG	CGGTG	C TT	GACG	GTGT	GGC	G											264
(2)	INFOR	MATI	ON F	OR S	EQ I	D NO	:164	:									

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1171 base pairs

(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

TAGTCGGCGA CGATGACGTC GCGGTCCAGG CCGACCGCTT CAAGCACCAG CGCGACCACG 60 AAGCCGGTGC GATCCTTACC CGCGAAGCAG TGGGTGAGCA CCGGGCGTCC GGCGGCAAGC 120 AGTGTGACGA CACGATGTAG CGCGCGCTGT GCTCCATTGC GCGTTGGGAA TTGGCGATAC 180 TCGTCGGTCA TGTAGCGGGT GGCCGCGTCA TTTATCGACT GGCTGGATTC GCCGGACTCG 240 CCGTTGGACC CGTCATTGGT TAGCAGCCTC TTGAATGCGG TTTCGTGCGG CGCTGAGTCG 300 TCGGCGTCAT CATCGGCGAG GTCGGGGAAC GGCAGCAGGT GGACGTCGAT GCCGTCCGGA 360 ACCCGTCCTG GACCGCGGC GGCAACCTCC CGGGACGACC GCAGGTCGGC AACGTCGGTG 420 ATCCCCAGCC GGCGCAGCGT TGCCCCTCGT GCCGAATTCG GCACGAGGCT GGCGAGCCAC 480 CGGGCATCAC CAAGCAACGC TTGCCCAGTA CGGATCGTCA CTTCCGCATC CGGCAGACCA ATCTCCTCGC CGCCCATCGT CAGATCCCGC TCGTGCGTTG ACAAGAACGG CCGCAGATGT 600 GCCAGCGGGT ATCGGAGATT GAACCGCGCA CGCAGTTCTT CAATCGCTGC GCGCTGCCGC 660 ACTATTGGCA CTTTCCGGCG GTCGCGGTAT TCAGCAAGCA TGCGAGTCTC GACGAACTCG 720 CCCCACGTAA CCCACGGCGT AGCTCCCGGC GTGACGCGGA GGATCGGCGG GTGATCTTTG 780 CCGCCACGCT CGTAGCCGTT GATCCACCGC TTCGCGGTGC CGGCGGGGAG GCCGATCAGC 840 TTATCGACCT CGGCGTATGC CGACGGCAAG CTGGGCGCGT TCGTCGAGGT CAAGAACTCC 900 ACCATCGGCA CCGGCACCAA GGTGCCGCAC CTGACCTACG TCGGCGACGC CGACATCGGC 960 GAGTACAGCA ACATCGGCGC CTCCAGCGTG TTCGTCAACT ACGACGGTAC GTCCAAACGG 1020 CGCACCACCG TCGGTTCGCA CGTACGGACC GGGTCCGACA CCATGTTCGT GGCCCCAGTA 1080 ACCATCGGCG ACGGCGCGTA TACCGGGGCC GGCACAGTGG TGCGGGAGGA TGTCCCGCCG GGGGCGCTGG CAGTGTCGGC GGGTCCGCAA C 1171

(2) INFORMATION FOR SEQ ID NO:165:

- (i) SEQUENCE CHARACTERISTIGS:
 - (A) LENGTH: 227 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

GCAAAGGCGG CACCGGCGGG GCCGGCATGA ACAGCCTCGA CCCGCTGCTA GCCGCCCAAG	60
ACGGCGGCCA AGGCGGCACC GGCGGCACCG GCGCCACCAGCT	120
TCACCCAAGG CGCCGACGGC AACGCCGGCA ACGGCGGTGA CGGCGGGGTC GGCGGCAACG	180
GCGGAAACGG CGGAAACGGC GCAGACAACA CCACCACCGC CGCCGCC	227
(2) INFORMATION FOR SEQ ID NO:166:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 304 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:	
CCTCGCCACC ATGGGCGGGC AGGGCGGTAG CGGTGGCGCC GGCTCTACCC CAGGCGCCAA	60
GGGCGCCCAC GGCTTCACTC CAACCAGCGG CGGCGACGGC GGCGACGGCG GCAACGGCGG	120
CAACTCCCAA GTGGTCGGCG GCAACGGCGG CGACGGCGGC AATGGCGGCA ACGGCGGCAG	180
CGCCGGCACG GGCGGCAACG GCGGCCGCGG CGGCGACGGC GCGTTTGGTG GCATGAGTGC	240
CAACGCCACC AACCCTGGTG AAAACGGGCC AAACGGTAAC CCCGGCGGCA ACGGTGGCGC	300
CGGC	304
(2) INFORMATION FOR SEQ ID NO:167:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1439 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:	
GTGGGACGCT GCCGAGGCTG TATAACAAGG ACAACATCGA CCAGCGCCGG CTCGGTGAGC	60
TGATCGACCT ATTTAACAGT GCGCGCTTCA GCCGGCAGGG CGAGCACCGC GCCCGGGATC	120
TGATGGGTGA GGTCTACGAA TACTTCCTCG GCAATTTCGC TCGCGCGGAA GGGAAGCGGG	180
GTGGCGAGTT CTTTACCCCG CCCAGCGTGG TCAAGGTGAT CGTGGAGGTG CTGGAGCCGT	240
CGAGTGGGCG GGTGTATGAC CCGTGCTGCG GTTCCGGAGG CATGTTTGTG CAGACCGAGA	300
AGTTCATCTA CGAACACGAC GGCGATCCGA AGGATGTCTC GATCTATGGC CAGGAAAGCA	360
TTGAGGAGAC CTGGCGGATG GCGAAGATGA ACCTCGCCAT CCACGGCATC GACAACAAGG	420

GGCTCGGCG	C CCGATGGAGT	GATACCTTCG	CCCGCGACCA	GCACCCGGA	GTGCAGATGG	480
ACTACGTGAT	GGCCAATCCG	CCGTTCAACA	TCAAAGACTG	GGCCCGCAAC	GAGGAAGACC	540
CACGCTGGC	CTTCGGTGTT	CCGCCGCCA	ATAACGCCAA	CTACGCATGG	ATTCAGCACA	600
TCCTGTACAA	CTTGGCGCCG	GGAGGTCGGG	CGGGCGTGGT	GATGGCCAAC	GGGTCGATGT	660
CGTCGAACTC	CAACGGCAAG	GGGGATATTC	GCGCGCAAAT	CGTGGAGGCG	GATTTGGTTT	720
CCTGCATGGT	CGCGTTACCC	ACCCAGCTGT	TCCGCAGCAC	CGGAATCCCG	GTGTGCCTGT	780
GGTTTTTCGC	CAAAAACAAG	GCGGCAGGTA	AGCAAGGGTC	TATCAACCGG	TGCGGGCAGG	840
TGCTGTTCAT	CGACGCTCGT	GAACTGGGCG	ACCTAGTGGA	CCGGGCCGAG	CGGGCGCTGA	900
CCAACGAGGA	GATCGTCCGC	ATCGGGGATA	CCTTCCACGC	GAGCACGACC	ACCGGCAACG	960
CCGGCTCCGG	TGGTGCCGGC	GGTAATGGGG	GCACTGGCCT	CAACGGCGCG	GGCGGTGCTG	1020
GCGGGGCCGG	CGGCAACGCG	GGTGTCGCCG	GCGTGTCCTT	CGGCAACGCT	GTGGGCGGCG	1080
ACGGCGGCAA	CGGCGGCAAC	GGCGGCCACG	GCGGCGACGG	CACGACGGGC	GGCGCCGGCG	1140
GCAAGGGCGG	CAACGGCAGC	AGCGGTGCCG	CCAGCGGCTC	AGGCGTCGTC	AACGTCACCG	1200
CCGGCCACGG	CGGCAACGGC	GGCAATGGCG	GCAACGGCGG	CAACGGCTCC	GCGGGCGCCG	1260
GCGGCCAGGG	CGGTGCCGGC	GGCAGCGCCG	GCAACGGCGG	CCACGGCGGC	GGTGCCACCG	1320
GCGGCGCCAG	CGGCAAGGGC	GGCAACGGCA	CCAGCGGTGC	CGCCAGCGGC	TCAGGCGTCA	1380
TCAACGTCAC	CGCCGGCCAC	GGCGGCAACG	GCGGCAATGG	CCGCAACGGC	GGCAACGGC	1439

(2) INFORMATION FOR SEQ ID NO:168:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 329 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

GGGCCGGCGG	GGCCGGATTT	TCTCGTGCCT	TGATTGTCGC	TGGGGATAAC	GGCGGTGATG	60
GTGGTAACGG	CGGGATGGGC	GGGGCTGGCG	GGGCTGGCGG	CCCCGGCGGG	GCCGGCGGCC	120
TGATCAGCCT	GCTGGGCGGC	CAAGGCGCCG	-GCGGGGCCGG	CGGGACCGGC	GGGGCCGGCG	180
GTGTTGGCGG	TGACGGCGGG	GCCGGCGGCC	CCGGCAACCA	GGCCTTCAAC	GCAGGTGCCG	240
GCGGGGCCGG	CGGCCTGATC	AGCCTGCTGG	GCGGCCAAGG	CGCCGGCGGG	GCCGGCGGGA	300
CCGGCGGGC	CGGCGGTGTT	GGCGGTGAC				329

(2) INFORMATION FOR SEQ ID NO:169:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 80 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:	
GCAACGGTGG CAACGGCGGC ACCAGCACGA CCGTGGGGAT GGCCGGAGGT AACTGTGGTG	60
CCGCCGGGCT GATCGGCAAC	80
(2) INFORMATION FOR SEQ ID NO:170:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 392 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:	
GGGCTGTGTC GCACTCACAC CGCCGCATTC GGCGACGTTG GCCGCCCAAT ATCCAGCTCA	60
AGGCCTACTA CTTACCGTCG GAGGACCGCC GCATCAAGGT GCGGGTCAGC GCCCAAGGAA	120
TCAAGGTCAT CGACCGCGAC GGGCATCGAG GCCGTCGTCG CGCGGCTCGG GCAGGATCCG	180
CCCCGGCGCA CTTCGCGCGC CAAGCGGGCT CATCGCTCCG AACGGCGGCG ATCCTGTGAG	240
CACAACTGAT GGCGCGCAAC GAGATTCGTC CAATTGTCAA GCCGTGTTCG ACCGCAGGGA	300
CCGGTTATAC GTATGTCAAC CTATGTCACT CGCAAGAACC GGCATAACGA TCCCGTGATC	360
CGCCGACAGC CCACGAGTGC AAGACCGTTA CA	392
(2) INFORMATION FOR SEQ ID NO:171:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 535 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:	
ACCGGCGCCA CCGGCGGCAC CGGGTTCGCC GGTGGCGCCG GCGGGGCCGG CGGGCAGGGC	60
GGTATCAGCG GTGCCGGCGG CACCAACGGC TCTGGTGGCG CTGGCGGCAC CGGCGGACAA	120
GGCGGCGCCG GGGGCCGGC GCCGATAACC CCACCGGCAT CGGCGGCGCC	180

GGCGGCACCG GCGGCACCGG C	GGAGCGGCC	GGAGCCGGCG	GGGCCGGTGG	CGCCATCGGT	240
ACCGGCGGCA CCGGCGGCGC G	GTGGGCAGC	GTCGGTAACG	CCGGGATCGG	CGGTACCGGC	300
GGTACGGGTG GTGTCGGTGG TO	GCTGGTGGT	GCAGGTGCGG	CTGCGGCCGC	TGGCAGCAGC	360
GCTACCGGTG GCGCCGGGTT CC	GCCGGCGGC	GCCGGCGGAG	AAGGCGGACC	GGGCGGCAAC	420
AGCGGTGTGG GCGGCACCAA CO	GCTCCGGC	GGCGCCGGCG	GTGCAGGCGG	CAAGGGCGGC	480
ACCGGAGGTG CCGGCGGGTC CC	GGCGCGGAC	AACCCCACCG	GTGCTGGTTT	CGCCG	53 5
(2) INFORMATION FOR SEQ	ID NO:172	:			
(i) SEQUENCE CHARAC	TOTTOTOTOTO				
(A) LENGTH: 69					
(B) TYPE: nucl		113			
(C) STRANDEDNE					
(D) TOPOLOGY:		-			
(D) TOPOLOGY:	ilnear				

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

CCGACGTCGC CGGGGCGATA CGGGGGTCAC CGACTACTAC ATCATCCGCA CCGAGAATCG 60 GCCGCTGCTG CAACCGCTGC GGGCGGTGCC GGTCATCGGA GATCCGCTGG CCGACCTGAT 120 CCAGCCGAAC CTGAAGGTGA TCGTCAACCT GGGCTACGGC GACCCGAACT ACGGCTACTC 180 GACGAGCTAC GCCGATGTGC GAACGCCGTT CGGGCTGTGG CCGAACGTGC CGCCTCAGGT CATCGCCGAT GCCCTGGCCG CCGGAACACA AGAAGGCATC CTTGACTTCA CGGCCGACCT 300 GCAGGCGCTG TCCGCGCAAC CGCTCACGCT CCCGCAGATC CAGCTGCCGC AACCCGCCGA 360 TCTGGTGGCC GCGGTGGCCG CCGCACCGAC GCCGGCCGAG GTGGTGAACA CGCTCGCCAG 420 GATCATCTCA ACCAACTACG CCGTCCTGCT GCCCACCGTG GACATCGCCC TCGCCTGGTC 480 ACCACCCTGC CGCTGTACAC CACCCAACTG TTCGTCAGGC AACTCGCTGC GGGCAATCTG ATCAACGCGA TCGGCTATCC CCTGGCGGCC ACCGTAGGTT TAGGCACGAT CGATAGCGGG 600 CGGCGTGGAA TTGCTCACCC TCCTCGCGGC GGCCTCGGAC ACCGTTCGAA ACATCGAGGG 660 CCTCGTCACC TAACGGATTC CCGACGGCAT 690

(2) INFORMATION FOR SEQ ID NO:173:

- (i) SEQUENCE CHARACTERISTIGS:
 - (A) LENGTH: 407 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

ACGGTGACGG CGGTACTGGC GGCGGCCACG GCGGCAACGG CGGGAATCCC GGGTGGCTCT	60
TGGGCACAGC CGGGGGTGGC GGCAACGGTG GCGCCGGCAG CACCGGTACT GCAGGTGGCG	120
GCTCTGGGGG CACCGGCGGC GACGGCGGGA CCGGCGGGCG TGGCGGCCTG TTAATGGGCG	180
CCGGCGCGG CGGGCACGGT GGCACTGGCG GCGCGGGCGG TGCCGGTGTC GACGGTGGCG	240
GCGCCGGCGG GCCGGCGGG GCCGGCGGCA ACGGCGGCGC CGGGGGTCAA GCCGCCCTGC	300
TGTTCGGGCG CGGCGGCACC GGCGGAGCCG GCGGCTACGG CGGCGATGGC GGTGGCGGCG	360
GTGACGGCTT CGACGGCACG ATGGCCGGCC TGGGTGGTAC CGGTGGC	407
(2) INFORMATION FOR SEQ ID NO:174:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 468 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:	
GATCGGTCAG CGCATCGCCC TCGGCGGCAA GCGATTCCGC GGTCTCACCG AAGAACATCG	60
TGCACGCGGC GGCGCGGACC AGCCCGCTGC GCTGCGGCGC GTCGAACGCC TCCAGCAGGC	120
ACAGCCAGTC CTTGGCGGCC TGCGAGGCGA ACACGTCGGT GTCACCGGTG TAGATCGCCG	180
GGATGCCCGC CTCCGCCAAC GCATTCCGGC ACGCCCGCGC GTCTTTGTGA TGCTCGACGA	240
TCACCGCGAT GTCTGCGGCC ACCACGGGCC GCCCGGCGAA GGTGGCCCCG CTGGCCAGTA	300
GCGCCGCGAC GTCGGCGGCC AGGTCGTCGG GGATGTGCCG GCGCAGCGCT CCGGCGCGAC	360
GCCCGAAAAA CGACCCCTCA CCCAGCTGGG TCCCGCTGGC ATATCCCTTG CCGTCCTGGG	420
CGATATTGGA CGCGCATGCC CCGACCGCGT ACAGGCCGGC CACCACCG	468
(2) INFORMATION FOR SEQ ID NO:175:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 219 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:	
GGTGGTAACG GCGGCCAGGG TGGCATCGGC GGCGCCGGCG AGAGAGGCGC CGACGGCGCC	60

GGCCCCAATG CTAACGGCGC AAACGGCGAG AACGGCGGTA GCGGTGGTAA CGGTGGCGAC 120

GGCGGCGCCG GCGGCAATGG CGGCGCGGGC GGCAACGCGC AGGCGGCCGG GTACACCGAC	180
GGCGCCACGG GCACCGGCGG CGACGGCGGC AACGGCGGC	219
(2) INFORMATION FOR SEQ ID NO:176:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 494 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:	
TAGCTCCGGC GAGGGCGGCA AGGGCGGCGA CGGTGGCCAC GGCGGTGACG GCGTCGGCGG	60
CAACAGTTCC GTCACCCAAG GCGGCAGCGG CGGTGGCGGC GGCGCCGGCGG	120
CAGCGGCTTT TTCGGCGGCA AGGGCGGCTT CGGCGGCGAC GGCGGTCAGG GCGGCCCCAA	180
CGGCGGCGGT ACCGTCGGCA CCGTGGCCGG TGGCGGCGGC AACGGCGGTG TCGGCGGCCG	240
GGGCGGCGAC GGCGTCTTTG CCGGTGCCGG CGGCCAGGGC GGCCTCGGTG GGCAGGGCGG	300
CAATGGCGGC GGCTCCACCG GCGGCAACGG CGGCCTTGGC GGCGCGGGGG GTGGCGGAGG	360
CAACGCCCCG GCTCGTGCCG AATCCGGGCT GACCATGGAC AGCGCGGCCA AGTTCGCTGC	420
CATCGCATCA GGCGCGTACT GCCCCGAACA CCTGGAACAT CACCCGAGTT AGCGGGGCGC	480
ATTTCCTGAT CACC	494
(2) INFORMATION FOR SEQ ID NO:177:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 220 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:	
GGGCCGGTGG TGCCGCGGGC CAGCTCTTCA GCGCCGGAGG CGCGGCGGGT GCCGTTGGGG	60
TTGGCGGCAC CGGCGGCCAG GGTGGGGCTG GCGGTGCCGG AGCGGCCGGC GCCGACGCCC	120
CCGCCAGCAC AGGTCTAACC GGTGGTACCG GGTTCGCTGG CGGGGCCGGC GGCGTCGGCG	180
GCCAGAGCGG CAACGCCATT GCCGGCGGCA TCAACGGCTC	220
(2) INFORMATION FOR SEQ ID NO:178:	
(i) SEQUENCE CHARACTERISTICS:	

(A) LENGTH: 388 base pairs (B) TYPE: nucleic acid

WO 99/42076 PCT/US99/03268

(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:	
ATGGCGGCAA CGGGGGCCCC GGCGGTGCTG GCGGGGCCGG CGACTACAAT TTCCAACGGC	60
GGGCAGGGTG GTGCCGGCGG CCAAGGCGGC CAAGGCGGCC TGGGCGGGC AAGCACCACC	120
TGATCGGCCT AGCCGCACCC GGGAAAGCCG ATCCAACAGG CGACGATGCC GCCTTCCTTG	180
CCGCGTTGGA CCAGGCCGGC ATCACCTACG CTGACCCAGG CCACGCCATA ACGGCCGCCA	240
AGGCGATGTG TGGGCTGTGT GCTAACGGCG TAACAGGTCT ACAGCTGGTC GCGGACCTGC	
GGGACTACAA TCCCGGGCTG ACCATGGACA GCGCGGCCAA GTTCGCTGCC ATCGCATCAG	360
GCGCGTACTG CCCCGAACAC CTGGAACA	388
(2) INFORMATION FOR SEQ ID NO:179:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 400 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:	
GCAAAGGCGG CACCGGCGGG GCCGGCATGA ACAGCCTCGA CCCGCTGCTA GCCGCCCAAG	60
ACGGCGGCCA AGGCGGCACC GGCGCCACCG GCGCCACCCC CGGCGCCGGC GGCACCAGCT	120
TCACCCAAGG CGCCGACGGC AACGCCGGCA ACGGCGGTGA CGGCGGGGGTC GGCGGCAACG	180
GCGGAAACGG CGGAAACGGC GCAGACAACA CCACCACCGC CGCCGCCGGC ACCACAGGCG	240
GCGACGGCGG GGCCGGCGGG GCCGGCGGAAC CCGGCGGAACC GCCGGCACCG	300
GCACCGGCGG CCAACAAGGC AACGGCGGCA ACGGCGGCAC CGGCGGCAAA GGCGGCACCG	360
GCGGCGACGG TGCACTCTCA GGCAGCACCG GTGGTGCCGG	400
(2) INFORMATION FOR SEQ ID NO:180:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 538 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:	
GGCAACGGCG GCAACGGCGG CATCGCCGGC ATTGGGCGGC AACGGCGTTC CGGGACGGGC	6.0

WO 99/42076 PCT/US99/03268

AGCGGCAACG GCGGCCAACG GCGCCAGCGG CGGCAACGGC GCCAACGCCG GCATGGGCGG	120
CAACAGCGGC ACCGGCAGCG GCGACGGCGG TGCCGGCGGG AACGGCGGCG CGGCGGCAC	180
GGGCGGCACC GGCGGCGACG GCGGCCTCAC CGGTACTGGC GGCACCGGCG GCAGCGGTGG	240
CACCGGCGGT GACGGCGGTA ACGGCGGCAA CGGAGCAGAT AACACCGCAA ACATGACTGC	300
GCAGGCGGGC GGTGACGGTG GCAACGGCGG CGACGGTGGC TTCGGCGGCG GGGCCGGGGC	360
CGGCGGCGGT GGCTTGACCG CTGGCGCCAA CGGCACCGGC GGGCAAGGCG GCGCCGGCGG	420
CGATGGCGGC AACGGGGCCA TCGGCGGCCA CGGCCCACTC ACTGACGACC CCGGCGGCAA	480
CGGGGGCACC GGCGGCAACG GCGGCACCGG CGGCACCGGC GGCGCGGGCA TCGGCAGC	538
(2) INFORMATION FOR SEQ ID NO:181:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 239 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:	
GGGCCGGTGG TGCCGCGGGC CAGCTCTTCA GCGCCGGAGG CGCGGCGGGT GCCGTTGGGG	60
TTGGCGGCAC CGGCGGCCAG GGTGGGGCTG GCGGTGCCGG AGCGGCCGGC GCCGACGCCC	120
CCGCCAGCAC AGGTCTAACC GGTGGTACCG GGTTCGCTGG CGGGGCCGGC GGCGTCGGCG	180
GCCACGGCGG CAACGCCATT GCCGGCGGCA TCAACGGCTC CGGTGGTGCC GGCGGCACC	239
(2) INFORMATION FOR SEQ ID NO:182:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 985 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:	
AGCAGCGCTA CCGGTGGCGC CGGGTTCGCC GGCGGCGCCG GCGGAGAAGG CGGAGCGGGC	60
GGCAACAGCG GTGTGGGCGG CACCAACGGC TCCGGCGGCG CCGGCGGTGC AGGCGGCAAG	120
GGCGGCACCG GAGGTGCCGG CGGGTCCGGC GCGGACAACC CCACCGGTGC TGGTTTCGCC	180
GGTGGCGCCG GCGCACAGG TGGCGCGGCC GGCGCCGGCG GGCGACCGGT	240
ACCGGCGGCA CCGGCGGCGT TGTCGGCGCC ACCGGTAGTG CAGGCATCGG CGGGGCCGGC	300
GGCCGCGGCG GTGACGGCGG CGATGGGGCC AGCGGTCTCG GCCTGGGCCT CTCCGGCTTT	360

Checececo						
					CGGCATCAAC	420
GGGGCCGGCG	GGGCCGGCGG	CAACGGCGGC	GACGGCGGG	ACGGCGCAAC	CGGTGCCGCA	480
GGTCTCGGCG	ACAACGGCGG	GGTCGGCGGT	GACGGTGGGG	CCGGTGGCGC	CGCCGGCAAC	540
GGCGGCAACG	CGGGCGTCGG	CCTGACAGCC	AAGGCCGGCG	ACGGCGGCGC	CGCGGGCAAT	600
					CAACGGCGGC	660
					CACCACCTGA	720
		AAAGCCGATC				780
		ACCTACGCTG				840
		AACGGCGTAA				
		ATGGACAGCG				900
CGTACTGCCC				CGCIGCCATC	GCATCAGGCG	960
		CARCA				985

(2) INFORMATION FOR SEQ ID NO:183:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2138 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

CGGCACGAGG ATCGGTACCC CGCGGCATCG GCAGCTGCCG ATTCGCCGGG TTTCCCCACC 60 CGAGGAAAGC CGCTACCAGA TGGCGCTGCC GAAGTAGGGC GATCCGTTCG CGATGCCGGC 120 ATGAACGGGC GGCATCAAAT TAGTGCAGGA ACCTTTCAGT TTAGCGACGA TAATGGCTAT 180 AGCACTAAGG AGGATGATCC GATATGACGC AGTCGCAGAC CGTGACGGTG GATCAGCAAG 240 AGATTTTGAA CAGGGCCAAC GAGGTGGAGG CCCCGATGGC GGACCCACCG ACTGATGTCC 300 CCATCACACC GTGCGAACTC ACGGCGGCTA AAAACGCCGC CCAACAGCTG GTATTGTCCG CCGACAACAT GCGGGAATAC CTGGCGGCCG GTGCCAAAGA GCGGCAGCGT CTGGCGACCT 420 CGCTGCGCAA CGCGGCCAAG GCGTATGGCG AGGTTGATGA GGAGGCTGCG ACCGCGCTGG 480 ACAACGACGG CGAAGGAACT GTGCAGGCAG AATCGGCCGG GGCCGTCGGA GGGGACAGTT 540 CGGCCGAACT AACCGATACG CCGAGGGTGG CCACGGCCGG TGAACCCAAC TTCATGGATC 600 TCAAAGAAGC GGCAAGGAAG CTCGAAACGG GCGACCAAGG CGCATCGCTC GCGCACTTTG

CGGATGGGTG GAACACTTTC AACCTGACGC TGCAAGGCGA CGTCAAGCGG	TTCCGGGGGT	720
TTGACAACTG GGAAGGCGAT GCGGCTACCG CTTGCGAGGC TTCGCTCGAT	CAACAACGGC	780
AATGGATACT CCACATGGCC AAATTGAGCG CTGCGATGGC CAAGCAGGCT	CAATATGTCG	840
CGCAGCTGCA CGTGTGGGCT AGGCGGGAAC ATCCGACTTA TGAAGACATA	GTCGGGCTCG	900
AACGGCTTTA CGCGGAAAAC CCTTCGGCCC GCGACCAAAT TCTCCCGGTG	TACGCGGAGT	960
ATCAGCAGAG GTCGGAGAAG GTGCTGACCG AATACAACAA CAAGGCAGCC	CTGGAACCGG	1020
TAAACCCGCC GAAGCCTCCC CCCGCCATCA AGATCGACCC GCCCCCGCCT	CCGCAAGAGC	1080
AGGGATTGAT CCCTGGCTTC CTGATGCCGC CGTCTGACGG CTCCGGTGTG	ACTCCCGGTA	1140
CCGGGATGCC AGCCGCACCG ATGGTTCCGC CTACCGGATC GCCGGGTGGT	GGCCTCCCGG	1200
CTGACACGGC GGCGCAGCTG ACGTCGGCTG GGCGGGAAGC CGCAGCGCTG	TCGGGCGACG	1260
TGGCGGTCAA AGCGGCATCG CTCGGTGGCG GTGGAGGCGG CGGGGTGCCG	TCGGCGCCGT	1320
TGGGATCCGC GATCGGGGGC GCCGAATCGG TGCGGCCCGC TGGCGCTGGT	GACATTGCCG	1380
GCTTAGGCCA GGGAAGGGCC GGCGGCGGCG CCGCGCTGGG CGGCGGTGGC	ATGGGAATGC	1440
CGATGGGTGC CGCGCATCAG GGACAAGGG GCGCCAAGTC CAAGGGTTCT	CAGCAGGAAG	1500
ACGAGGCGCT CTACACCGAG GATCGGGCAT GGACCGAGGC CGTCATTGGT	AACCGTCGGC	1560
GCCAGGACAG TAAGGAGTCG AAGTGAGCAT GGACGAATTG GACCCGCATG	TCGCCCGGC	1620
GTTGACGCTG GCGGCGCGGT TTCAGTCGGC CCTAGACGGG ACGCTCAATC	AGATGAACAA	1680
CGGATCCTTC CGCGCCACCG ACGAAGCCGA GACCGTCGAA GTGACGATCA	ATGGGCACCA	1740
GTGGCTCACC GGCCTGCGCA TCGAAGATGG TTTGCTGAAG AAGCTGGGTG	CCGAGGCGGT	1800
GGCTCAGCGG GTCAACGAGG CGCTGCACAA TGCGCAGGCC GCGGCGTCCG	CGTATAACGA	1860
CGCGGCGGGC GAGCAGCTGA CCGCTGCGTT ATCGGCCATG TCCCGCGCGA	TGAACGAAGG	1920
AATGGCCTAA GCCCATTGTT GCGGTGGTAG CGACTACGCA CCGAATGAGC	GCCGCAATGC	1980
GGTCATTCAG CGCGCCCGAC ACGGCGTGAG TACGCATTGT CAATGTTTTG		2040
GCCGGGTTCG GAGGGCGCCA TAGTCCTGGT CGCCAATATT GCCGCAGCTA		2100
GGTTCGGTTA CGCTGGTTAA TTATGACGTC CGTTACCA		2139

(2) INFORMATION FOR SEQ ID NO:184:

⁽i) SEQUENCE CHARACTERISTICS:

⁽A) LENGTH: 460 amino acids

⁽B) TYPE: amino acid

- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:
- Met Thr Gln Ser Gln Thr Val Thr Val Asp Gln Gln Glu Ile Leu Asn 1 5 10 10
- Arg Ala Asn Glu Val Glu Ala Pro Met Ala Asp Pro Pro Thr Asp Val
- Pro Ile Thr Pro Cys Glu Leu Thr Ala Ala Lys Asn Ala Ala Gln Gln 35 40 45
- Leu Val Leu Ser Ala Asp Asn Met Arg Glu Tyr Leu Ala Ala Gly Ala 50 55 60
- Lys Glu Arg Gln Arg Leu Ala Thr Ser Leu Arg Asn Ala Ala Lys Ala 65 70 75 80
- Tyr Gly Glu Val Asp Glu Glu Ala Ala Thr Ala Leu Asp Asn Asp Gly 85 90 95
- Glu Gly Thr Val Gln Ala Glu Ser Ala Gly Ala Val Gly Gly Asp Ser
- Ser Ala Glu Leu Thr Asp Thr Pro Arg Val Ala Thr Ala Gly Glu Pro
- Asn Phe Met Asp Leu Lys Glu Ala Ala Arg Lys Leu Glu Thr Gly Asp
- Gln Gly Ala Ser Leu Ala His Phe Ala Asp Gly Trp Asn Thr Phe Asn 145 150 155 160
- Leu Thr Leu Gln Gly Asp Val Lys Arg Phe Arg Gly Phe Asp Asn Trp 165 170 175
- Glu Gly Asp Ala Ala Thr Ala Cys Glu Ala Ser Leu Asp Gln Gln Arg 180 185 190
- Gln Trp Ile Leu His Met Ala Lys Leu Ser Ala Ala Met Ala Lys Gln
 195 200 205
- Ala Gln Tyr Val Ala Gln Leu His Val Trp Ala Arg Arg Glu His Pro 210 215 220
- Thr Tyr Glu Asp Ile Val Gly Leu Glu Arg Leu Tyr Ala Glu Asn Pro 225 230 235 240
- Ser Ala Arg Asp Gln Ile Leu Pro Val Tyr Ala Glu Tyr Gln Gln Arg
- Ser Glu Lys Val Leu Thr Glu Tyr Asn Asn Lys Ala Ala Leu Glu Pro

WO 99/42076 PCT/US99/03268

175

Val	Asn	Pro	Pro	Lys	Pro	Pro	Pro	Ala	Tle	Luc	т1.	3	n	_	_
		275							+10	Lys	TTE	Asp	PLO	Pro	Pro
							280					285			

- Pro Pro Gln Glu Gln Gly Leu Ile Pro Gly Phe Leu Met Pro Pro Ser
- Asp Gly Ser Gly Val Thr Pro Gly Thr Gly Met Pro Ala Ala Pro Met 305 310 315
- Val Pro Pro Thr Gly Ser Pro Gly Gly Gly Leu Pro Ala Asp Thr Ala 325 330 335
- Ala Gln Leu Thr Ser Ala Gly Arg Glu Ala Ala Ala Leu Ser Gly Asp 340 345 350
- Val Ala Val Lys Ala Ala Ser Leu Gly Gly Gly Gly Gly Gly Val
 355 360 365
- Pro Ser Ala Pro Leu Gly Ser Ala Ile Gly Gly Ala Glu Ser Val Arg 370 375 380
- Pro Ala Gly Ala Gly Asp Ile Ala Gly Leu Gly Gln Gly Arg Ala Gly 385 390 395
- Gly Gly Ala Ala Leu Gly Gly Gly Gly Met Gly Met Pro Met Gly Ala
 405
 415
- Ala His Gln Gly Gln Gly Gly Ala Lys Ser Lys Gly Ser Gln Gln Glu
- Asp Glu Ala Leu Tyr Thr Glu Asp Arg Ala Trp Thr Glu Ala Val Ile
- Gly Asn Arg Arg Arg Gln Asp Ser Lys Glu Ser Lys 450 455

(2) INFORMATION FOR SEQ ID NO:185:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 277 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

- Ala Gly Asn Val Thr Ser Ala Ser Gly Pro His Arg Phe Gly Ala Pro 1 5 10 15
- Asp Arg Gly Ser Gln Arg Arg Arg His Pro Ala Ala Ser Thr Ala
 20 25 30
- Thr Glu Arg Cys Arg Phe Asp Arg His Val Ala Arg Gln Arg Cys Gly 35 40 45

- Phe Pro Pro Ser Arg Arg Gln Leu Arg Arg Arg Val Ser Arg Glu Ala 50 55 60
- Thr Thr Arg Arg Ser Gly Arg Arg Asn His Arg Cys Gly Trp His Pro
- Gly Thr Gly Ser His Thr Gly Ala Val Arg Arg Arg His Gln Glu Ala
- Arg Asp Gln Ser Leu Leu Leu Arg Arg Gly Arg Val Asp Leu Asp 100 105 110
- Gly Gly Gly Arg Leu Arg Arg Val Tyr Arg Phe Gln Gly Cys Leu Val
- Val Val Phe Gly Gln His Leu Leu Arg Pro Leu Leu Ile Leu Arg Val
- His Arg Glu Asn Leu Val Ala Gly Arg Arg Val Phe Arg Val Lys Pro 150 155 160
- Phe Glu Pro Asp Tyr Val Phe Ile Ser Arg Met Phe Pro Pro Ser Pro 165 170 175
- His Val Gln Leu Arg Asp Ile Leu Ser Leu Leu Gly His Arg Ser Ala 180 185 190
- Gln Phe Gly His Val Glu Tyr Pro Leu Pro Leu Leu Ile Glu Arg Ser
- Leu Ala Ser Gly Ser Arg Ile Ala Phe Pro Val Val Lys Pro Pro Glu
- Pro Leu Asp Val Ala Leu Gln Arg Gln Val Glu Ser Val Pro Pro Ile 225 230 235 240
- Arg Lys Val Arg Glu Arg Cys Ala Leu Val Ala Arg Phe Glu Leu Pro
- Cys Arg Phe Phe Glu Ile His Glu Val Gly Phe Thr Gly Arg Gly His

Pro Arg Arg Ile Gly 275

- (2) INFORMATION FOR SEQ ID NO:186:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 192 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

- Arg Val Ala Ala Ser Phe Ile Asp Trp Leu Asp Ser Pro Asp Ser Pro 1 5 10 10
- Leu Asp Pro Ser Leu Val Ser Ser Leu Leu Asn Ala Val Ser Cys Gly
 20 25 30
- Ala Glu Ser Ser Ala Ser Ser Ser Ala Arg Ser Gly Asn Gly Ser Arg
 35 40 45
- Trp Thr Ser Met Pro Ser Gly Thr Arg Pro Gly Pro Arg Arg Ala Thr 50 55 60
- Ser Arg Asp Asp Arg Arg Ser Ala Thr Ser Val Ile Pro Ser Arg Arg 65 70 75
- Ser Val Ala Pro Arg Ala Glu Phe Gly Thr Arg Leu Ala Ser His Arg 85 90 95
- Ala Ser Pro Ser Asn Ala Cys Pro Val Arg Ile Val Thr Ser Ala Ser
- Gly Arg Pro Ile Ser Ser Pro Pro Ile Val Arg Ser Arg Ser Cys Val
- Asp Lys Asn Gly Arg Arg Cys Ala Ser Gly Tyr Arg Arg Leu Asn Arg
- Ala Arg Ser Ser Ser Ile Ala Ala Arg Cys Arg Thr Ile Gly Thr Phe
- Arg Arg Ser Arg Tyr Ser Ala Ser Met Arg Val Ser Thr Asn Ser Pro
- His Val Thr His Gly Val Ala Pro Gly Val Thr Arg Arg Ile Gly Gly 180 185 190
- (2) INFORMATION FOR SEQ ID NO:187:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 196 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:
 - Gln Glu Arg Pro Gln Met Cys Gln Arg Val Ser Glu Ile Glu Pro Arg

 1 10 15
 - Thr Gln Phe Phe Asn Arg Cys Ala Leu Pro His Tyr Trp His Phe Pro
 - Ala Val Ala Val Phe Ser Lys His Ala Ser Leu Asp Glu Leu Ala Pro

WO 99/42076 PCT/US99/03268

178

40 **4**5

Arg Asn Pro Arg Arg Ser Ser Arg Arg Asp Ala Glu Asp Arg Arg Val

Ile Phe Ala Ala Thr Leu Val Ala Val Asp Pro Pro Leu Arg Gly Ala 70 75 80

Gly Gly Glu Ala Asp Gln Leu Ile Asp Leu Gly Val Cys Arg Arg Gln 85 90 95

Ala Gly Arg Val Arg Arg Gly Gln Glu Leu His His Arg His Arg His 100 105 110

Gln Gly Ala Ala Pro Asp Leu Arg Arg Arg Arg Arg His Arg Arg Val

Gln Gln His Arg Arg Leu Gln Arg Val Arg Gln Leu Arg Arg Tyr Val 130 135 140

Gln Thr Ala His His Arg Arg Phe Ala Arg Thr Asp Arg Val Arg His

His Val Arg Gly Pro Ser Asn His Arg Arg Arg Arg Val Tyr Arg Gly

Arg His Ser Gly Ala Gly Gly Cys Pro Ala Gly Gly Ala Gly Ser Val

Gly Gly Ser Ala 195

(2) INFORMATION FOR SEQ ID NO:188:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 311 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

Val Arg Cys Gly Thr Leu Val Pro Val Pro Met Val Glu Phe Leu Thr 1 5 10 10

Ser Thr Asn Ala Pro Ser Leu Pro Ser Ala Tyr Ala Glu Val Asp Lys 20 25 30

Leu Ile Gly Leu Pro Ala Gly Thr Ala Lys Arg Trp Ile Asn Gly Tyr 35 40 45

Glu Arg Gly Gly Lys Asp His Pro Pro Ile Leu Arg Val Thr Pro Gly
50 55 60

Ala Thr Pro Trp Val Thr Trp Gly Glu Phe Val Glu Thr Arg Met Leu

WO 99/42076 PCT/US99/03268

179

65	;					70					75					80
Al	a G	lu	Ту	r Ar	g As 85	p Ar	g Ar	g Ly	s Va	1 Pr	o Il	e Va	l Ar	g Glr	1 Arg 95	J Ala
Al	a I	le	Gli	u Gl 10	u Le O	u Ar	g Al	a Ar	g Ph 10:	e Ası 5	ı Leı	ı Arg	ŢŢ	Pro		Ala
Hi	s L	eu	Arg	y Pr	o Ph	e Lei	ı Se	Th:	r His	s Glu	ı Arg	J Asi	Lev 125		Met	Gly
Gl	y G. 1.	lu 30	Glu	ıIl	e Gly	y Lei	135	Ası	Ala	a Glu	ı Val	Thr	: Ile	Arg	Thr	Gly
Gl: 145	1 A.	la	Leu	Le	ı Gly	/ Asp 150	Ala	Arg	Trp	Leu	Ala 155	Ser	Leu	Val	Pro	Asn 160
Ser	. Al	.a	Arg	Gly	/ Ala 169	Thr	Leu	Arg	Arg	Leu 170	Gly	Ile	Thr	Asp	Val 175	Ala
Asp	Le	u	Arg	Ser 180	Ser	Arg	Glu	Val	Ala 185	Arg	Arg	Gly	Pro	Gly 190	Arg	Val
Pro	As	g	Gly 195	Ile	Asp	Val	His	Leu 200	Leu	Pro	Phe	Pro	Asp 205	Leu	Ala	Asp
Asp	As 21	р . 0	Ala	Asp	Asp	Ser	Ala 215	Pro	His	Glu	Thr	Ala 220	Phe	Lys	Arg	Leu
Leu 225	Th	r j	Asn	Asp	Gly	Ser 230	Asn	Gly	Glu	Ser	Gly 235	Glu	Ser	Ser	Gln	Ser 240
Ile	Ası	1 A	/z p	Ala	Ala 245	Thr	Arg	Tyr	Met	Thr 250	Asp	Glu	Tyr	Arg	Gln 255	Phe
Pro	Thi	: <i>p</i>	urg	Asn 260	Gly	Ala	Gln	Arg	Ala 265	Leu	His	Arg	Val	Val 270	Thr	Leu
Leu	Ala	. A	1 a 75	Gly	Arg	Pro	Val	Leu 280	Thr	His	Cys	Phe	Ala 285	Gly	Lys	Asp
Arg	Thr 290	G	ly	Phe	Val	Val	Ala 295	Leu	Val	Leu	Glu	Ala 300	Val	Gly	Leu .	Asp

(2) INFORMATION FOR SEQ ID NO:189:

(i) SEQUENCE CHARACTERISTICS:

Arg Asp Val Ile Val Ala Asp

- (A) LENGTH: 2072 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:189:

CTCGTGCCGA TTCGGCACGA GCTGAGCAGC CCAAGGGGCC GTTCGGCGAA GTCATCGAGG	60
CATTCGCCGA CGGGCTGGCC GGCAAGGGTA AGCAAATCAA CACCACGCTG AACAGCCTGT	120
CGCAGGCGTT GAACGCCTTG AATGAGGGCC GCGGCGACTT CTTCGCGGTG GTACGCAGCC	180
TGGCGCTATT CGTCAACGCG CTACATCAGG ACGACCAACA GTTCGTCGCG TTGAACAAGA	240
ACCTTGCGGA GTTCACCGAC AGGTTGACCC ACTCCGATGC GGACCTGTCG AACGCCATCC	300
AGCAATTCGA CAGCTTGCTC GCCGTCGCGC GCCCGTTCTT CGCCAAGAAC CGCGAGGTGC	360
TGACGCATGA CGTCAATAAT CTCGCGACCG TGACCACCAC GTTGCTGCAG CCCGATCCGT	420
TGGATGGGTT GGAGACCGTC CTGCACATCT TCCCGACGCT GGCGGCGAAC ATTAACCAGC	480
TTTACCATCC GACACACGGT GGCGTGGTGT CGCTTTCCGC GTTCACGAAT TTCGCCAACC	540
CGATGGAGTT CATCTGCAGC TCGATTCAGG CGGGTAGCCG GCTCGGTTAT CAAGAGTCGG	600
CCGAACTCTG TGCGCAGTAT CTGGCGCCAG TCCTCGATGC GATCAAGTTC AACTACTTTC	660
CGTTCGGCCT GAACGTGGCC AGCACCGCCT CGACACTGCC TAAAGAGATC GCGTACTCCG	720
AGCCCCGCTT GCAGCCGCCC AACGGGTACA AGGACACCAC GGTGCCCGGC ATCTGGGTGC	780
CGGATACGCC GTTGTCACAC CGCAACACGC AGCCCGGTTG GGTGGTGGCA CCCGGGATGC	840
AAGGGGTTCA GGTGGGACCG ATCACGCAGG GTTTGCTGAC GCCGGAGTCC CTGGCCGAAC	900
TCATGGGTGG TCCCGATATC GCCCCTCCGT CGTCAGGGCT GCAAACCCCG CCCGGACCCC	960
CGAATGCGTA CGACGAGTAC CCCGTGCTGC CGCCGATCGG TTTACAGGCC CCACAGGTGC	1020
CGATACCACC GCCGCCTCCT GGGCCCGACG TAATCCCGGG TCCGGTGCCA CCGGTCTTGG	1080
CGGCGATCGT GTTCCCAAGA GATCGCCCGG CAGCGTCGGA AAACTTCGAC TACATGGGCC	1140
TCTTGTTGCT GTCGCCGGGC CTGGCGACCT TCCTGTTCGG GGTGTCATCT AGCCCCGCCC	1200
GTGGAACGAT GGCCGATCGG CACGTGTTGA TACCGGCGAT CACCGGCCTG GCGTTGATCG	1260
CGGCATTCGT CGCACATTCG TGGTACCGCA CAGAACATCC GCTCATAGAC ATGCGCTTGT	1320
TCCAGAACCG AGCGGTCGCG CAGGCCAACA TGACGATGAC GGTGCTCTCC CTCGGGCTGT	1380
TTGGCTCCTT CTTGCTGCTC CCGAGCTACC TCCAGCAAGT GTTGCACCAA TCACCGATGC	1440
AATCGGGGGT GCATATCATC CCACAGGGCC TCGGTGCCAT GCTGGCGATG CCGATCGCCG	1500
GAGCGATGAT GGACCGACGG GGACCGGCCA AGATCGTGCT GGTTGGGATC ATGCTGATCG	1560
CTGCGGGGTT GGGCACCTTC GCCTTTGGTG TCGCGCGGCA AGCGGACTAC TTACCCATTC	1620

TGCCGACCGG	GCTCCC3 3 TC	> magaa				
					CCACTGTCCG	1680
GGGCGGCAGT	GCAGACCCTG	GCCCCACATC	AGATCGCTCG	CGGTTCGACG	CTGATCAGCG	1740
TCAACCAGCA	GGTGGGCGGT	TCGATAGGGA	CCGCACTGAT	GTCGGTGCTG	CTCACCTACC	1800
AGTTCAATCA	CAGCGAAATC	ATCGCTACTG	CAAAGAAAGT	CGCACTGACC	CCAGAGAGTG	1860
GCGCCGGGCG	GGGGGCGCG	GTTGACCCTT	CCTCGCTACC	GCGCCAAACC	AACTTCGCGG	1920
CCCAACTGCT	GCATGACCTT	TCGCACGCCT	ACGCGGTGGT	ATTCGTGATA	GCGACCGCGC	1980
TAGTGGTCTC	GACGCTGATC	CCCGCGGCAT	TCCTGCCGAA	ACAGCAGGCT	AGTCATCGAA	2040
GAGCACCGTT	GCTATCCGCA	TGACGTCTGC	TT			2072

(2) INFORMATION FOR SEQ ID NO:190:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1923 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

TCACCCCGGA GAAGTCGTTC GTCGACGACC TGGACATCGA CTCGCTGTCG ATGGTCGAGA 60 TCGCCGTGCA GACCGAGGAC AAGTACGGCG TCAAGATCCC CGACGAGGAC CTCGCCGGTC 120 TGCGTACCGT CGGTGACGTT GTCGCCTACA TCCAGAAGCT CGAGGAAGAA AACCCGGAGG 180 CGGCTCAGGC GTTGCGCGCG AAGATTGAGT CGGAGAACCC CGATGCGGCA CGAGCAGATC 240 GGTGCGTTTC ACCCACATCG CAAGCTCGAG ACGCCCGTCG TCCTCTTGCA CGCTCAGCCA 300 GGTTGGCGTG TCGCCGCCTT CCAGCAAGTG TTCCCACCAC ACGAAGGGAC CCTCGCGAAA 360 GGTGACTGAT CCGCGGACCA CATAGTCGAT GCCACCGTGG CTGACAATTG CGCCGGGTCC 420 GAGTTGGCGG GGGCCGAATT GCGGCATTGC GTCGAAGGCC AGCGGATCCC GGCGCCCGCC 480 CGGCGTGGCT GGTGTTTTGG GCCGCCGGAT GGCCACGACG AGAACGACGA TGGCGGCGAT 540 GAACAGCGCC ACGGCAATCA CGACCAGCAG ATTTCCCACG CATACCCTCT CGTACCGCTG 600 CGCCGCGGTT GGTCGATCGG TCGCATATCG ATGGCGCCGT TTAACGTAAC AGCTTTCGCG 660 GGACCGGGGG TCACAACGGG CGAGTTGTCC GGCCGGGAAC CCGGCAGGTC TCGGCCGCGG 720 TCACCCCAGC TCACTGGTGC ACCATCCGGG TGTCGGTGAG CGTGCAACTC AAACACACTC 780 AACGGCAACG GTTTCTCAGG TCACCAGCTC AACCTCGACC CGCAATCGCT CGTACGTTTC 840

	AGGTCGCGAG					900
	TAGGTTGCGC					960
GCGATATGAG	CAGGCATCCA	GTGCCAGGTA	GTTGCTGGAG	GTGATGTCCG	CCAAGTAGGC	1020
GTGGACGGCA	ACAGGGGCAA	TACGATGCGG	CGGTGGTAGC	CGGGTCAAGA	CCGAATAGGT	1080
TTCCACAGCC	GCGTGCGCGA	TCAGATGGAC	GCCACGGTTG	AGCGCGCGCA	CGGCGGCCTC	1140
GTGCCCTTCG	TGCCAGGTCG	CGAATCCGGC	AACCAGCACG	CTGGTGTCTG	GTGCGATCAC	1200
CGCCGTGTGC	GATCGAGCGT	TTCCCGAACG	ATTTCGTCGG	TCAACGGGGG	CAGGGGACGT	1260
TCTGGCCGTG	CGACGAGAAC	CGAGCCTTCC	CGAACGAGTT	CGACACCGGT	CGGGGCCGGC	1320
TCAATCTCGA	TGCGCCCATC	GCGCTCGGTG	ATCTCCACCT	GGTCGTTCCC	GCGCAAGCCA	1380
AGGCGCTCGC	GAATCCGCTT	GGGAATCACC	AGACGTCCTG	CGACATCGAT	GGTTGTTCGC	1440
ATGGTAGGAA	ATTTACCATC	GCACGTTCCA	TAGGCGTGTC	CTGCGCGGGA	TGTCGGGACG	1500
ATCCGCTAGC	GTATCGAACG	ATTGTTTCGG	AAATGGCTGA	GGGAGCGTGC	GGTGCGGGTG	1560
ATGGGTGTCG	ATCCCGGGTT	GACCCGATGC	GGGCTGTCGC	TCATCGAGAG	TGGGCGTGGT	1620
CGGCAGCTCA	CCGCGCTGGA	TGTCGACGTG	GTGCGCACAC	CGTCGGATGC	GGCCTTGGCG	1680
CAGCGCCTGT	TGGCCATCAG	CGATGCCGTC	GAGCACTGGC	TGGACACCCA	TCATCCGGAG	1740
GTGGTGGCTA	TCGAACGGGT	GTTCTCTCAG (CTCAACGTGA	CCACGGTGAT (GGGCACCGCG	1800
CAGGCCGGCG	GCGTGATCGC (CCTGGCGGCG (ECCAAACGTG (GTGTCGACGT (GCATTTCCAT	1860
ACCCCCAGCG 2	AGGTCAAGGC (GCGGTCACT (GCAACGGTT (CCGCAGACAA (GCTCAGGTC	1920
ACC						1923

(2) INFORMATION FOR SEQ ID NO:191:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1055 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

CTGGCGTGCC	AGTGTCACCG	GCGATATGAC	GTCGGCATTC	AATTTCGCGG	CCCCGCCGGA	60
CCCGTCGCCA	CCCAATCTGG	ACCACCCGGT	CCGTCAATTG	CCGAAGGTCG	CCAAGTGCGT	120
GCCCAATGTG	GTGCTGGGTT	TCTTGAACGA	AGGCCTGCCG	TATCGGGTGC	CCTACCCCCA	180
AACAACGCCA	GTCCAGGAAT	CCGGTCCCGC	GCGGCCGATT	CCCAGCGGCA	TCTGCTAGCC	240

GGGG33						
			G GTCGAAACCC			300
GCTCATGGC	GCGAAATTAG	AAAACCCGGC	ATATTGTCCG	CGGATTGTCA	TACGATGCTG	360
			GTGTGGATGT			420
			CCTTTGCCGC			480
			TCGGTTCCGC			540
			GCGCCGGAGG			600
			GCGGCCATGG			660
			GAACCCCGCT			720
			GGCTAGGAGA			
			CCGCGGCCGG			780
			TCAACAAGGA			840
						900
			ACATGGCGGG (960
AGACGGTGGA	GGACGAGGCT	CGCCGGATGT	GGGCGTCCGC	GCAAAACATC 1	rcgggngcgg	1020
GCTGGAGTGG	CATGGCCGAG	GCGACCTCGC	TAGAC			1055

(2) INFORMATION FOR SEQ ID NO:192:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 359 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:192:

CCGCCTCGT GTTGCCATAC TCCGCCGCGG CCGCCTCGAC CGCACTGGCC GTGGCGTGTG 60
TCCGGGGCTGA CCACCGGGAT CGCCGAACCA TCCGAGATCA CCTCGCAATG ATCCACCTCG 120
CGCAGCTGGT CACCCAGCCA CCGGGCGGT TGCGACAGCG CCTGCATCAC CTTGGTATAG 180
CCGTCGCGCC CCAGCCGCAG GAAGTTGTAG TACTGGCCCA CCACCTGGTT ACCGGGACGG 240
GAGAAGTTCA GGGTGAAGGT CGGCATGTCG CCGCCGAGGT AGTTGACCCG GAAAACCAGA 300
TCCTCCGGCA GGTGCTCGGG CCCGCGCCAC ACGACAAACC CGACGCCGGG ATAGGTCAG 359

- (2) INFORMATION FOR SEQ ID NO:193:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 350 base pairs

(B) TYPE: nucleic acid

		(C)	STRA	NDED:	NESS : li	: si : si	ngle								
(x	:i) S					ION:	SEQ	ID N	10: 19	93 :					
AACGGG											GCAT	GAA	GTGC	TGGA	AG
GATGCA															
CGCAAA															
TGCGCG															
TGATCG													AACC	CAGT	GG
GTGGCC	rgga	AGAG	GTGC	TC T	ACGA	GCTG	T CT	CCGA'	TCGA	GGA	CTTC	TCC			
(2) INE	FORMA	MOIT	FOR	SEQ	ID.	NO:1	94:								
	() () ()	A) L: B) T C) S D) T	ENGT: YPE: TRANI DPOL	H: 6 ami; DEDNI DGY:	79 am no am ESS: line	∍ar	acio								
						ON: S									
Gl 1	u Gli	ı Pro	Lys	5 Gly	/ Pro	Phe	e Gly	Glu	Val	Ile	Glu	Ala	Phe	Ala 15	Ası
						' Lys		25					30		
						. Leu	40					45			
						Leu 55					60				
					, 0	Asn				75					80
				33		Asp			90					95	
						Arg		105					110		
						Asn	120					125			
Gln	Pro 130	Asp	Pro	Leu	qzA	Gly 135	Leu	Glu	Thr	Val	Leu 140	His	Ile	Phe	Pro

Th	r L	eu	Ala	a Al	a As	n Il 15	e As: O	n Gli	n Le	u Ty:	r Hi:	s Pro) Th	F His	s Gly	7 Gly 160
					10	5				170)				175	
				10	U				185	5				190	•	Ser
			4,5	•				200)				205			Lys
							215	•				220				Thr
	-					230	,				235					Asn 240
					443	,		. Pro		250					255	
				200				Pro	265					270		
		•						Ile 280					285			
							295	Gly				300				
						310		Pro			315					320
					323			Gln		330					335	
				340				Ile	345					350		
		_	-					Asp 360					365			
							3/5	Leu				380				
						390		Ala			395					400
					103			Gly		410					415	
				120					425					430		
FIIC	-111	A.	-11 £	-rrg	мта	va⊥	Ala	Gln	Ala	Asn	Met	Thr i	Met '	Thr '	Val	Leu

		43	5				440)				445	;		
Sea	£ Let 450	ı Gly	y Let	ı Phe	≘ Gly	Ser 455	Phe	Leu	Leu	. Leu	Pro		Туг	Leu	Gln
Glr 469	ı Val	. Le	ı His	Glr	Ser 470	Pro	Met	Gln	Ser	Gly 475	Val	His	Ile	Ile	Pro 480
Gln	Gly	' Leu	Gly	/ Ala 485	Met	Leu	Ala	Met	Pro 490	Ile	Ala	Gly	Ala	Met 495	Met
Asp	Arg	Arg	Gly 500	Pro	Ala	Lys	Ile	Val 505	Leu	Val	Gly	Ile	Me t 510	Leu	Ile
Ala	Ala	Gly 515	Leu	Gly	Thr	Phe	Ala 520	Phe	Gly	Val	Ala	Arg 525	Gln	Ala	Asp
Tyr	Leu 530	Pro	Ile	Leu	Pro	Thr 535	Gly	Leu	Ala	Ile	Met 540	Gly	Met	Gly	Met
Gly 54 5	Cys	Ser	Met	Met	Pro 550	Leu	Ser	Gly	Ala	Ala 555	Val	Gln	Thr	Leu	Ala 560
Pro	His	Gln	Ile	Ala 565	Arg	Gly	Ser	Thr	Leu 570	Ile	Ser	Val	Asn	Gln 575	Gln
Val	Gly	Gly	Ser 580	Ile	Gly	Thr	Ala	Leu 585	Met	Ser	Val	Leu	Leu 590	Thr	Tyr
Gln	Phe	Asn 595	His	Ser	Glu	Ile	Ile 600	Ala	Thr	Ala	Lys	Lys 605	Val	Ala	Leu
Thr	Pro 610	Glu	Ser	Gly	Ala	Gly 615	Arg	Gly	Ala	Ala	Val 620	Asp	Pro	Ser	Ser
Leu 625	Pro	Arg	Gln	Thr	Asn 630	Phe	Ala	Ala	Gln	Leu 635	Leu	His	Asp		Ser 640
His	Ala	Tyr	Ala	Val 645	Val	Phe	Val	Ile	Ala 650	Thr	Ala	Leu	Val	Val 655	Ser

Thr Leu Ile Pro Ala Ala Phe Leu Pro Lys Gln Gln Ala Ser His Arg

665

Arg Ala Pro Leu Leu Ser Ala 675

660

(2) INFORMATION FOR SEQ ID NO:195:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 120 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:
- Thr Pro Glu Lys Ser Phe Val Asp Asp Leu Asp Ile Asp Ser Leu Ser 1 5 10 10 15
- Met Val Glu Ile Ala Val Gln Thr Glu Asp Lys Tyr Gly Val Lys Ile
- Pro Asp Glu Asp Leu Ala Gly Leu Arg Thr Val Gly Asp Val Val Ala
 35 40 45
- Tyr Ile Gln Lys Leu Glu Glu Glu Asn Pro Glu Ala Ala Gln Ala Leu
 50 55 60
- Arg Ala Lys Ile Glu Ser Glu Asn Pro Asp Ala Ala Arg Ala Asp Arg 65 70 75 80
- Cys Val Ser Pro Thr Ser Gln Ala Arg Asp Ala Arg Arg Pro Leu Ala 85 90 95
- Arg Ser Ala Arg Leu Ala Cys Arg Arg Leu Pro Ala Ser Val Pro Thr
- Thr Arg Arg Asp Pro Arg Glu Arg
- (2) INFORMATION FOR SEQ ID NO:196:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 89 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:
 - Leu Ala Cys Gln Cys His Arg Arg Tyr Asp Val Gly Ile Gln Phe Arg

 1 10 15
 - Gly Pro Ala Gly Pro Val Ala Thr Gln Ser Gly Pro Pro Gly Pro Ser 20 25 30
 - Ile Ala Glu Gly Arg Gln Val Arg Ala Gln Cys Gly Ala Gly Phe Leu
 35 40 45
 - Glu Arg Arg Pro Ala Val Ser Gly Ala Leu Pro Pro Asn Asn Ala Ser 50 55 60
 - 9ro Gly Ile Arg Ser Arg Ala Ala Asp Ser Gln Arg His Leu Leu Ala 65 70 75
 - Gly Asp Gly Ser Asp Val Thr Val Gly
- (2) INFORMATION FOR SEQ ID NO:197:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 119 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:
- Ala Ser Leu Leu Ala Tyr Ser Ala Ala Ala Ala Ser Thr Ala Leu Ala 1 5 10 15
- Val Ala Cys Val Arg Ala Asp His Arg Asp Arg Arg Thr Ile Arg Asp 20 25 30
- His Leu Ala Met Ile His Leu Ala Gln Leu Val Thr Gln Pro Pro Gly
 35 40 45
- Gly Val Arg Gln Arg Leu His His Leu Gly Ile Ala Val Ala Pro Gln 50 55 60
- Pro Gln Glu Val Val Leu Ala His His Leu Val Thr Gly Thr Gly 70 75 80
- Glu Val Gln Gly Glu Gly Arg His Val Ala Ala Glu Val Val Asp Pro
- Glu Asn Gln Ile Leu Arg Gln Val Leu Gly Pro Ala Pro His Asp Lys
 100 105 110
- Pro Asp Ala Gly Ile Gly Gln 115
- (2) INFORMATION FOR SEQ ID NO:198:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 116 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:
 - Arg Ala Arg Gly His Arg Ser Ser Lys Gly Ser Arg Trp Ser His Glu

 10 15
 - Val Leu Glu Gly Cys Ile Leu Ala Asp Ser Arg Gln Ser Lys Thr Ala
 - Ala Ser Pro Ser Pro Ser Arg Pro Gln Ser Ser Ser Asn Asn Ser Val
 - Pro Gly Ala Pro Asn Arg Val Ser Phe Ala Lys Leu Arg Glu Pro Leu 50 55 60

Glu 65	Val	Pro	Gly	Leu	Leu 70	Asp	Val	Gln	Thr	Asp 75	Ser	Phe	Glu	Trp	Leu 80
Ile	Gly	Ser	Pro	Arg 85	Trp	Arg	Glu	Ser	Ala 90	Ala	Glu	Arg	Gly	Asp 95	Val
Asn	Pro	Val	Gly 100	Gly	Leu	Glu	Glu	Val 105	Leu	Tyr	Glu	Leu	Ser	Pro	Ile
Glu	Asp	Phe 115	Ser												

(2) INFORMATION FOR SEQ ID NO:199:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 811 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:

TOOMS OF SE						
			GTGGATGCCG			60
GTGAAAGCCG	CCGACGTGTT	CGCCGCATTC	GGGGAGAACA	TCGAACTGCT	CAAAAGGCTG	120
GTGCGGGCCG	CCATCGATCG	GGTCGCCGAC	GAGCGCACGT	GCACGCACTG	TCAACACCAC	180
GCCGGTGTTC	CGTTGCCGTT	CGAGCTGCCA	TGAGGGTGCT	GCTGACCGGC	GCGGCCGGCT	240
TCATCGGGTC	GCGCGTGGAT	GCGGCGTTAC	GGGCTGCGGG	TCACGACGTG	GTGGGCGTCG	300
ACGCGCTGCT	GCCCGCCGCG	CACGGGCCAA	ACCCGGTGCT	GCCACCGGGC	TGCCAGCGGG	360
			CGTTGTTGGC			420
ACCAGGCCGC	CATGGTGGGT	GCCGGCGTCA	ACGCCGCCGA	CGCACCCGCC	TATGGCGGCC	480
ACAACGATTT						540
			GGCAGGGGCG			600
GACCGGTCGA						660
GTTGCCCGGG						720
GCCCGCGCAG						780
GGAAACGAAT (1000010010	
						811

(2) INFORMATION FOR SEQ ID NO:200:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 966 base pairs
 - (B) TYPE: nucleic acid

(C)	STRANDEDNESS:	single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:

CTCCCCCCC						
	T GTGGCCGAGO					60
GACTTTGTG	G TCCCGGTGGC	GGGATAGAGC	ACCTGTCGGC	GTTGGTCAGC	GTCACCCGTT	120
GCTCGGACG	C CGAACCCATG	CTTTCAACGT	AGCCTGTCGG	TCACACAAGT	CGCGAGCGTA	180
ACGTCACGG	T CAAATATCGC	GTGGAATTTC	GCCGTGACGT	TCCGCTCGCG	GACAATCAAG	240
	C TTACATGCGA					300
	A GGTGTACAGC					360
	A CGGACGCATC					420
	1 TCTTGCCCGG					480
	CGCCAGCTTG					540
	CGACGTGGAC					600
	GAAATCGTAT					660
	TCGCCCCACC					720
	GCTCGGCTAC					
	AGAAGCCGCT					780
	TCACCCGATG					840
						900
CATCCT	GCATCCCACC	CGIICCCGGA	AACCGGTTCC (GCCGGGTCGG	CTCATCGCTT	960
Casa CC i						966

(2) INFORMATION FOR SEQ ID NO:201:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2367 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

CCGCACCGCC	GGCAATACCG	CCAGCGCCAC	CGTTACCGCC	GTTTGCGCCG	TTGCCCCCGT	6 0
TGCCGCCCGT	CCCGCCGGCC	CCGCCGATGG	AGTTCTCATC	GCCAAAAGTA	CTGGCGTTGC	120
CACCGGAGCC	GCCGTTGCCG	CCGTCACCGC	CAGCCCCGCC	GACTCCACCG	GCCCCACCGA	180

CTCCGCCGCT GCCACCGTTG CCGCCGTTGC CGATCAACAT GCCGCTGGCG CC	CACCCTTGC 240
CACCCACGCC ACCGGCTCCG CCCACCCCGC CGACACCAAG CGAGCTGCCG CC	EGGAGCCAC 300
CATCACCACC TACGCCACCG ACCGCCCAGA CACCAGCGAC CGGGTCTTCG TG	BAAACGTCG 360
CGGTGCCACC ACCGCCGCCG TTACCGCCAA CCCCACCGGC AACGCCGGCG CC	GCCATCCC 420
CGCCGGCCCC GGCGTTGCCG CCGTTGCCGAA CAACAACCCG CC	GGCGCCGC 480
CGTTGCCGCC CGCGCCGCCG GTCCCGCCGG CGCCGCCGAC GCCAAGGCCG CT	GCCGCCCT 540
TGCCGCCATC ACCACCCTTG CCGCCGACCA CATCGGGTTC TGCCTCGGGG TC	TGGGCTGT 600
CAAACCTCGC GATGCCAGCG TTGCCGCCGC TTCCCCCGGG CCCCCCGTG GC	GCCGTCAC 560
CACCGATACC ACCCGCGCCA CCGGCGCCAC CGTTGCCGCC ATCACCGAAT AG	CAACCCGC 720
CGGCGCCACC ATTGCCGCCA GCTCCCCCTG CGCCACCGTC GGCGCCGGAG GC	GGCACTGG 780
CAGCCCCGTT ACCACCGAAA CCGCCGCTAC CACCGGTAGA GGTGGCAGTG GCC	SATGTGTA 840
CGAAAGCGCC GCCTCCGGCG CCGCCGCTAC CACCCCCACT GCCGGCGGCT AC	ACCGTCGG 900
ACCCGTTGCC ACCATCACCG CCAAAGGCGC TCGCAATGTC GCCCTGCGCG ACT	rccgccgt 960
CGCCGCCGTT GCCGCCGCCG CCACCGGCAG CGGCGGTACC GCCGTCACCA CCC	GGCACCGC 1020
CGGTGGCCTT GCCCGAGCCT GCCGTCGCGG TGGCACCGTC GCCGCCGGTG CCA	ACCGGTCG 1080
GCGTGCCGGC AGTGCCATGG CCGCCCGTGC CGCCGTCGCC GCCGGTTTGA TCF	ACCGATGC 1140
CGGACACATC TGCCGGGCTG TCCCCGGTGC TGGCCGCGGG GCCGGGCGTG GGA	ATTGACCC 1200
CGTTTGCCCC GGCGAGGCCG GCGCCGCCGG TACCACCGGC GCCGCCATGG CCG	BAACAGCC 1260
CGGCGTTGCC GCCGTTACCG CCCGCACCCC CGATGCCTGC GGCCACGCTG GTG	CCGCCGA 1320
CACCGCCGTT GCCGCCGTTG CCCCACAACC ACCCCCCGTT CCCACCGGCA CCG	CCGGCCG 1380
CGCCGGTACC ACCGGCCCCG CCGTTGCCGC CGTTGCCGAT CAACCCGGCC GCG	CCTCCGC 1440
TGCCGCCGGT TTGACCGAAC CCGCCAGCCG CGCCGTTGCC ACCGTTGCCA AAC	AGCAACC 1500
CGCCGGCCGC GCCAGGCTGC CCGGGTGCCG TCCCGTCGGC GCCGTTTCCG ATC	AACGGGC 1560
GCCCCAAAAG CGCCTCGGTG GGCGCATTCA CCGCACCCAG CAGACTCCGC TCA	ACAGCGG 1620
CTTCAGTGCT GGCATACCGA CCCGCGGCCG CAGTCAACGC CTGCACAAAC TGC	TCGTGAA 1680
ACGCTGCCAC CTGTACGCTG AGCGCCTGAT ACTGCCGAGC ATGGGCCCCG AAC	AACCCCG 1740
CAATCGCCGC CGACACTTCA TCGGCAGCCG CAGCCACCAC TTCCGTCGTC GGG	ATCGCCG 1800
CGGCCGCATT AGCCGCGCTC ACCTGCGAAC CAATAGTCGA TAAATCCAAA GCC	GCAGTTG 1860

CCAGCAGCTG	CGGCGTCGCG	ATCACCAAGG	ACACCTCGCA	CCTCCGGATA	CCCCATATCG	1920
CCGCACCGTG	TCCCCAGCGG	CCACGTGACC	TTTGGTCGCT	GGCTGGCGGC	CCTGACTATG	1980
					AGTTGAAGAC	2040
					GTCCGCGGCG	2100
					AGTAGAGCCA	2160
CCTCAAGCTC						2220
					TCGGCCCAGG	2280
			TTCGAAGGGT	GGCTAGCGAC	GAGTTACAGC	2340
CCGCCTGGTA						2367
(2) INFORMAT	TION FOR SEC	Q ID NO:202	:			
(<i>)</i> (E (C	QUENCE CHARMAN LENGTH: [] Discrepance of the control of the contr	376 amino a ino acid NESS:	: cids			

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:202:

Gln Pro Ala Gly Ala Thr Ile Ala Ala Ser Ser Pro Cys Ala Thr Val

Gly Ala Gly Gly Gly Thr Gly Ser Pro Val Thr Thr Glu Thr Ala Ala
20 25 30

Thr Thr Gly Arg Gly Gly Ser Gly Asp Val Tyr Glu Ser Ala Ala Ser 35 40 45

Gly Ala Ala Ala Thr Thr Pro Thr Ala Gly Gly Tyr Thr Val Gly Pro
50 55 60

Val Ala Thr Ile Thr Ala Lys Gly Ala Arg Asn Val Ala Leu Arg Asp
65 70 75 80

Ser Ala Val Ala Ala Val Ala Ala Ala Thr Gly Ser Gly Gly Thr

Ala Val Thr Thr Gly Thr Ala Gly Gly Leu Ala Arg Ala Cys Arg Arg 100 105 110

Gly Gly Thr Val Ala Ala Gly Ala Thr Gly Arg Arg Ala Gly Ser Ala
115 120 125

Met Ala Ala Arg Ala Ala Val Ala Ala Gly Leu Ile Thr Asp Ala Gly
130 135 140

- His Ile Cys Arg Ala Val Pro Gly Ala Gly Arg Gly Ala Gly Arg Gly 145 150 155 160
- Ile Asp Pro Val Cys Pro Gly Glu Ala Gly Ala Ala Gly Thr Thr Gly 165 170 175
- Ala Ala Met Ala Glu Gln Pro Gly Val Ala Ala Val Thr Ala Arg Thr 180 185 190
- Pro Asp Ala Cys Gly His Ala Gly Ala Ala Asp Thr Ala Val Ala Ala
 195 200 205
- Val Ala Pro Gln Pro Pro Pro Val Pro Thr Gly Thr Ala Gly Arg Ala
 210 215 220
- Gly Thr Thr Gly Pro Ala Val Ala Ala Val Ala Asp Gln Pro Gly Arg
 225 230 235 240
- Ala Ser Ala Ala Gly Leu Thr Glu Pro Ala Ser Arg Ala Val Ala 245 250 255
- Thr Val Ala Lys Gln Gln Pro Ala Gly Arg Ala Arg Leu Pro Gly Cys 260 265 270
- Arg Pro Val Gly Ala Val Ser Asp Gln Arg Ala Pro Gln Lys Arg Leu 275 280 285
- Gly Gly Arg Ile His Arg Thr Gln Gln Thr Pro Leu Asn Ser Gly Phe
 290 295 300
- Ser Ala Gly Ile Pro Thr Arg Gly Arg Ser Gln Arg Leu His Lys Leu 305 310 315 320
- Leu Val Lys Arg Cys His Leu Tyr Ala Glu Arg Leu Ile Leu Pro Ser 325 330 335
- Met Gly Pro Glu Gln Pro Arg Asn Arg Arg Arg His Phe Ile Gly Ser 340 345 350
- Arg Ser His His Phe Arg Arg Arg Asp Arg Arg Gly Arg Ile Ser Arg
- Ala His Leu Arg Thr Asn Ser Arg 370 375
- (2) INFORMATION FOR SEQ ID NO:203:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2852 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:203:

GGCCAAAACG CCCCGGCGAT CGCGGCCACC GAGGCCGCCT ACGACCAGAT GTGGGCCCAG	60
GACGTGGCGG CGATGTTTGG CTACCATGCC GGGGCTTCGG CGGCCGTCTC GGCGTTGACA	120
CCGTTCGGCC AGGCGCTGCC GACCGTGGCG GGCGGCGGTG CGCTGGTCAG CGCGGCCGCG	180
GCTCAGGTGA CCACGCGGGT CTTCCGCAAC CTGGGCTTGG CGAACGTCCG CGAGGGCAAC	240
GTCCGCAACG GTAATGTCCG GAACTTCAAT CTCGGCTCGG	300
ATCGGCAGCG GCAACATCGG CAGCTCCAAC ATCGGGTTTG GCAACGTGGG TCCTGGGTTG	360
ACCGCAGCGC TGAACAACAT CGGTTTCGGC AACACCGGCA GCAACAACAT CGGGTTTGGC	420
AACACCGGCA GCAACAACAT CGGGTTCGGC AATACCGGAG ACGGCAACCG AGGTATCGGG	480
CTCACGGGTA GCGGTTTGTT GGGGTTCGGC GGCCTGAACT CGGGCACCGG CAACATCGGT	540
CTGTTCAACT CGGGCACCGG AAACGTCGGC ATCGGCAACT CGGGTACCGG GAACTGGGGC	600
ATTGGCAACT CGGGCAACAG CTACAACACC GGTTTTGGCA ACTCCGGCGA CGCCAACACG	660
GGCTTCTTCA ACTCCGGAAT AGCCAACACC GGCGTCGGCA ACGCCGGCAA CTACAACACC	720
GGTAGCTACA ACCCGGGCAA CAGCAATACC GGCGGCTTCA ACATGGGCCA GTACAACACG	780
GGCTACCTGA ACAGCGGCAA CTACAACACC GGCTTGGCAA ACTCCGGCAA TGTCAACACC	840
GGCGCCTTCA TTACTGGCAA CTTCAACAAC GGCTTCTTGT GGCGCGGCGA CCACCAAGGC	900
CTGATTTTCG GGAGCCCCGG CTTCTTCAAC TCGACCAGTG CGCCGTCGTC GGGATTCTTC	960
AACAGCGGTG CCGGTAGCGC GTCCGGCTTC CTGAACTCCG GTGCCAACAA TTCTGGCTTC	1020
TTCAACTCTT CGTCGGGGGC CATCGGTAAC TCCGGCCTGG CAAACGCGGG CGTGCTGGTA	1080
TCGGGCGTGA TCAACTCGGG CAACACCGTA TCGGGTTTGT TCAACATGAG CCTGGTGGCC	1140
ATCACAACGC CGGCCTTGAT CTCGGGCTTC TTCAACACCG GAAGCAACAT GTCGGGATTT	1200
TTCGGTGGCC CACCGGTCTT CAATCTCGGC CTGGCAAACC GGGGCGTCGT GAACATTCTC	1260
GGCAACGCCA ACATCGGCAA TTACAACATT CTCGGCAGCG GAAACGTCGG TGACTTCAAC	1320
ATCCTTGGCA GCGGCAACCT CGGCAGCCAA AACATCTTGG GCAGCGGCAA CGTCGGCAGC	1380
TTCAATATCG GCAGTGGAAA CATCGGAGTA TTCAATGTCG GTTCCGGAAG CCTGGGAAAC	1440
TACAACATCG GATCCGGAAA CCTCGGGATC TACAACATCG GTTTTGGAAA CGTCGGCGAC	1500
TACAACGTCG GCTTCGGGAA CGCGGGCGAC TTCAACCAAG GCTTTGCCAA CACCGGCAAC	1560
AACAACATCG GGTTCGCCAA CACCGGCAAC AACAACATCG GCATCGGGCT GTCCGGCGAC	1620

AACCAGCAG	G GCTTCAATAI	TGCTAGCGGC	TGGAACTCGG	GCACCGGCAA	CAGCGGCCTG	1680
TTCAATTCG	GCACCAATAA	CGTTGGCATC	TTCAACGCGG	GCACCGGAAA	CGTCGGCATC	1740
GCAAACTCGG	GCACCGGGAA	CTGGGGTATC	GGGAACCCGG	GTACCGACAA	TACCGGCATC	1800
CTCAATGCTG	GCAGCTACAA	CACGGGCATC	CTCAACGCCG	GCGACTTCAA	CACGGGCTTC	1860
TACAACACGG	GCAGCTACAA	CACCGGCGGC	TTCAACGTCG	GTAACACCAA	CACCGGCAAC	1920
TTCAACGTGG	GTGACACCAA	TACCGGCAGC	TATAACCCGG	GTGACACCAA	CACCGGCTTC	1980
TTCAATCCCG	GCAACGTCAA	TACCGGCGCT	TTCGACACGG	GCGACTTCAA	CAATGGCTTC	2040
TTGGTGGCGG	GCGATAACCA	GGGCCAGATT	GCCATCGATC	TCTCGGTCAC	CACTCCATTC	2100
ATCCCCATAA	ACGAGCAGAT	GGTCATTGAC	GTACACAACG	TAATGACCTT	CGGCGGCAAC	2160
ATGATCACGG	TCACCGAGGC	CTCGACCGTT	TTCCCCCAAA	CCTTCTATCT	GAGCGGTTTG	2220
TTCTTCTTCG	GCCCGGTCAA	TCTCAGCGCA	TCCACGCTGA	CCGTTCCGAC	GATCACCCTC	2280
ACCATCGGCG	GACCGACGGT	GACCGTCCCC	ATCAGCATTG	TCGGTGCTCT	GGAGAGCCGC	2340
ACGATTACCT	TCCTCAAGAT	CGATCCGGCG	CCGGGCATCG	GAAATTCGAC	CACCAACCCC	2400
TCGTCCGGCT	TCTTCAACTC	GGGCACCGGT	GGCACATCTG	GCTTCCAAAA	CGTCGGCGGC	2460
GGCAGTTCAG	GCGTCTGGAA	CAGTGGTTTG	AGCAGCGCGA	TAGGGAATTC	GGGTTTCCAG	2520
AACCTCGGCT	CGCTGCAGTC	AGGCTGGGCG	AACCTGGGCA	ACTCCGTATC	GGGCTTTTTC	2580
AACACCAGTA	CGGTGAACCT	CTCCACGCCG	GCCAATGTCT	CGGGCCTGAA	CAACATCGGC	2640
ACCAACCTGT	CCGGCGTGTT	CCGCGGTCCG	ACCGGGACGA	TTTTCAACGC	GGGCCTTGCC	2700
AACCTGGGCC	AGTTGAACAT	CGGCAGCGCC	TCGTGCCGAA	TTCGGCACGA	GTTAGATACG	2760
TTTCAACAA	TCATATCCGC	GTTTTGCGGC	AGTGCATCAG	ACGAATCGAA	CCCGGGAAGC	2820
STAAGCGAAT	AAACCGAATG	GCGGCCTGTC	AT			2852

(2) INFORMATION FOR SEQ ID NO:204:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 943 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:204:
- Gly Gln Asn Ala Pro Ala Ile Ala Ala Thr Glu Ala Ala Tyr Asp Gln

 1 10 15

WO 99/42076 PCT/US99/03268

196

- Met Trp Ala Gln Asp Val Ala Ala Met Phe Gly Tyr His Ala Gly Ala
 20 25 30
- Ser Ala Ala Val Ser Ala Leu Thr Pro Phe Gly Gln Ala Leu Pro Thr 35 40 45
- Val Ala Gly Gly Gly Ala Leu Val Ser Ala Ala Ala Ala Gln Val Thr
 50 55 60
- Thr Arg Val Phe Arg Asn Leu Gly Leu Ala Asn Val Arg Glu Gly Asn 75 80
- Val Arg Asn Gly Asn Val Arg Asn Phe Asn Leu Gly Ser Ala Asn Ile 85 90 95
- Gly Asn Gly Asn Ile Gly Ser Gly Asn Ile Gly Ser Ser Asn Ile Gly 100 105 110
- Phe Gly Asn Val Gly Pro Gly Leu Thr Ala Ala Leu Asn Asn Ile Gly 115 120 125
- Phe Gly Asn Thr Gly Ser Asn Asn Ile Gly Phe Gly Asn Thr Gly Ser 130 135 140
- Asn Asn Ile Gly Phe Gly Asn Thr Gly Asp Gly Asn Arg Gly Ile Gly
 145 150 155 160
- Leu Thr Gly Ser Gly Leu Leu Gly Phe Gly Gly Leu Asn Ser Gly Thr 165 to 170 to 175
- Gly Asn Ile Gly Leu Phe Asn Ser Gly Thr Gly Asn Val Gly Ile Gly 180 185 190
- Asn Ser Gly Thr Gly Asn Trp Gly Ile Gly Asn Ser Gly Asn Ser Tyr
 195 200 205
- Asn Thr Gly Phe Gly Asn Ser Gly Asp Ala Asn Thr Gly Phe Phe Asn 210 215 220
- Ser Gly Ile Ala Asn Thr Gly Val Gly Asn Ala Gly Asn Tyr Asn Thr 235 230 235 240
- Gly Ser Tyr Asn Pro Gly Asn Ser Asn Thr Gly Gly Phe Asn Met Gly 245 250 255
- Gln Tyr Asn Thr Gly Tyr Leu Asn Ser Gly Asn Tyr Asn Thr Gly Leu 260 265 270
- Ala Asn Ser Gly Asn Val Asn Thr Gly Ala Phe Ile Thr Gly Asn Phe
 275
 280
 285
- Asn Asn Gly Phe Leu Trp Arg Gly Asp His Gln Gly Leu Ile Phe Gly 290 295 300
- Ser Pro Gly Phe Phe Asn Ser Thr Ser Ala Pro Ser Ser Gly Phe Phe

د	05						31	0					31	5				320
A	sn (Ser	G1	y A.	la (31y 325	' Se	r Al	a Se	r G	ly	Phe 330	Le	u As	n Se	r Gl	y Al 33	a Asn 5
										34	.5					35	0	r Gly
Le	eu A	la	As. 35	n A] 5	La G	Sly	Va]	l Le	u Va 36	1 Se 0	r (Gly	Va:	l Il	e As 36		r Gl	y As n
								3/:	•					380	כ			r Pro
							390						395	i				/ Phe 400
					•	05					4	10					415	
					•					425	>					430	1	Gly
									440						445			Gly
								435						460				Gly
							4 /U						475			Leu		480
					10	_					4 9	90				Gly	495	
				-						505						Asp 510		
									520						525	Ala		
								233						540		Gln		
Phe 545	Asr	ıI	le	Ala	Se	r G 5	1y 50	Trp	Asn	Ser	G1	у 1 5	hr 55	Gly	Asn	Ser	Gly	Leu 560
					50.	•					57	0				Gly	575	
										585						Ile 590		
Pro	Gly	T) 59	hr 1 95	qzA	Asn	T	hr (Sly	Ile 600	Leu	As:	n A	la (Gly	Ser 605	Tyr	Asn	Thr

- Gly Ile Leu Asn Ala Gly Asp Phe Asn Thr Gly Phe Tyr Asn Thr Gly 610 615 620

 Ser Tyr Asn Thr Gly Gly Phe Asn Val Gly Asn Thr Asn Thr Gly Asn 625 630 635 640
- Phe Asn Val Gly Asp Thr Asn Thr Gly Ser Tyr Asn Pro Gly Asp Thr 645 650 655
- Asn Thr Gly Phe Phe Asn Pro Gly Asn Val Asn Thr Gly Ala Phe Asp 660 665 670
- Thr Gly Asp Phe Asn Asn Gly Phe Leu Val Ala Gly Asp Asn Gln Gly 675 680 685
- Gln Ile Ala Ile Asp Leu Ser Val Thr Thr Pro Phe Ile Pro Ile Asn 690 695 700
- Glu Gln Met Val Ile Asp Val His Asn Val Met Thr Phe Gly Gly Asn 705 710 715 720
- Met Ile Thr Val Thr Glu Ala Ser Thr Val Phe Pro Gln Thr Phe Tyr
 725 730 735
- Leu Ser Gly Leu Phe Phe Phe Gly Pro Val Asn Leu Ser Ala Ser Thr 740 745 750
- Leu Thr Val Pro Thr Ile Thr Leu Thr Ile Gly Gly Pro Thr Val Thr 755 760 765
- Val Pro Ile Ser Ile Val Gly Ala Leu Glu Ser Arg Thr Ile Thr Phe
 770 780
- Leu Lys Ile Asp Pro Ala Pro Gly Ile Gly Asn Ser Thr Thr Asn Pro 785 790 795 800
- Ser Ser Gly Phe Phe Asn Ser Gly Thr Gly Gly Thr Ser Gly Phe Gln 805 810
- Asn Val Gly Gly Ser Ser Gly Val Trp Asn Ser Gly Leu Ser Ser 820 825 830
- Ala Ile Gly Asn Ser Gly Phe Gln Asn Leu Gly Ser Leu Gln Ser Gly 835 840 845
- Trp Ala Asn Leu Gly Asn Ser Val Ser Gly Phe Phe Asn Thr Ser Thr 850 860
- Val Asn Leu Ser Thr Pro Ala Asn Val Ser Gly Leu Asn Asn Ile Gly 865 870 875 880
- Thr Asn Leu Ser Gly Val Phe Arg Gly Pro Thr Gly Thr Ile Phe Asn 885 890 895

WO 99/42076

	Ala	Gly	Leu	Ala 900	Asn	Leu	Gly	Gln	Leu 905	Asn	Ile	Gly	Ser	Ala 910	Ser	Cys	
	Arg	Ile	Arg 915	His	Glu	Leu	Asp	Thr 920	Val	Ser	Thr	Ile	Ile 925	Ser	Ala	Phe	
	Cys	Gly 930	Ser	Ala	Ser	Asp	Glu 935	Ser	Asn	Pro	Gly	Ser 940	Val	Ser	Glu		
(2)	INFO	RMATI	ON F	OR S	EQ I	D NO	:205	; :									
	(i)	(B) (C)	LENCE TYP STR TOP	GTH: E: n ANDE	53 ucle DNES	base ic a S: s	pai cid ingl	.rs									
	(xi)	SEQU	ENCE	DES	CRIP	TION	: SE	Q ID	NO:	205:							
GGA	TCCATA	AT GG	GCCA'	TCAT	CAT	CATC	ATC .	ACGT	GATC	GA C	ATCA	TCGG	G AC	С			53
(2)	INFOR	MATI	ON F	OR S	EQ I	D NO	:206	:									
	·(i)	(A) (B) (C)	ENCE LENC TYP! STRA	GTH: E: ni ANDEI	42 l lcle: ONES	base ic ad S: s:	pai: cid ingle	rs									
	(xi)	SEQUI	ENCE	DES	RIPT	rion:	: SE(O ID	NO:	206 :							
CCTC	SAATTC	A GG	CTC	GTT	GCGC	CCGG	CT (CATC	TGA	AC GA	4						42
(2)	INFOR	MATIC	ON FO	R SE	Q II	NO:	207:										
	(i) s	(B) (C)	ENCE LENG TYPE STRA TOPO	TH: : nu NDED	31 b clei NESS	ase .c ac : si	pair id ngle	s									
	(xi) S	SEQUE	NCE	DESC	RIPT	'ION :	SEQ	ID	NO : 2	07:							
GGAT	CCTGC	4 GGC	TCGA	AAC	CACC	GAGC	GG T	•									31
(2)	INFORM	tat io	n fo	R SE	Q ID	NO:	208:										
	(i) S	(A) (B) (C) (D)	LENG TYPE STRAI TOPOI	TH: : nu NDED: LOGY	31 b clei NESS : li	ase c ac : si near	pair id ngle										
	(xi) S	EQUE	NCE I	DESC	RIPT	ION:	SEO	TD 1	NT - 2	n a .							

CTCTGAATTC AGCGCTGGAA ATCGTCGCGA T	31
(2) INFORMATION FOR SEQ ID NO:209:	-
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 33 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:	
GGATCCAGCG CTGAGATGAA GACCGATGCC GCT	33
(2) INFORMATION FOR SEQ ID NO:210:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 38 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:	
GGATATCTGC AGAATTCAGG TTTAAAGCCC ATTTGCGA	38
(2) INFORMATION FOR SEQ ID NO:211:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 30 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:	
CCGCATGCGA GCCACGTGCC CACAACGGCC	20
(2) INFORMATION FOR SEQ ID NO:212:	30
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 37 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:	
CTTCATGGAA TTCTCAGGCC GGTAAGGTCC_GCTGCGG	37
(2) INFORMATION FOR SEQ ID NO:213:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 7676 base pairs	
(B) TYPE: nucleic acid	

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:213:

TGGCGAATGG GACGCGCCCT GTAGCGGCGC ATTAAGCGCG GCGGGTGTGG TGGTTACGCG 60 CAGCGTGACC GCTACACTTG CCAGCGCCCT AGCGCCCGCT CCTTTCGCTT TCTTCCCTTC 120 CTTTCTCGCC ACGTTCGCCG GCTTTCCCCG TCAAGCTCTA AATCGGGGGC TCCCTTTAGG 180 GTTCCGATTT AGTGCTTTAC GGCACCTCGA CCCCAAAAAA CTTGATTAGG GTGATGGTTC 240 ACGTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCCCT TTGACGTTGG AGTCCACGTT 300 CTTTAATAGT GGACTCTTGT TCCAAACTGG AACAACACTC AACCCTATCT CGGTCTATTC 360 TTTTGATTTA TAAGGGATTT TGCCGATTTC GGCCTATTGG TTAAAAAATG AGCTGATTTA 420 ACAAAAATTT AACGCGAATT TTAACAAAAT ATTAACGTTT ACAATTTCAG GTGGCACTTT 480 TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT CAAATATGTA 540 TCCGCTCATG AATTAATTCT TAGAAAAACT CATCGAGCAT CAAATGAAAC TGCAATTTAT 600 TCATATCAGG ATTATCAATA CCATATTTTT GAAAAAGCCG TTTCTGTAAT GAAGGAGAAA 660 ACTCACCGAG GCAGTTCCAT AGGATGGCAA GATCCTGGTA TCGGTCTGCG ATTCCGACTC 720 GTCCAACATC AATACAACCT ATTAATTTCC CCTCGTCAAA AATAAGGTTA TCAAGTGAGA 780 AATCACCATG AGTGACGACT GAATCCGGTG AGAATGGCAA AAGTTTATGC ATTTCTTTCC 840 AGACTTGTTC AACAGGCCAG CCATTACGCT CGTCATCAAA ATCACTCGCA TCAACCAAAC 900 CGTTATTCAT TCGTGATTGC GCCTGAGCGA GACGAAATAC GCGATCGCTG TTAAAAGGAC 960 AATTACAAAC AGGAATCGAA TGCAACCGGC GCAGGAACAC TGCCAGCGCA TCAACAATAT 1020 TTTCACCTGA ATCAGGATAT TCTTCTAATA CCTGGAATGC TGTTTTCCCG GGGATCGCAG 1080 TGGTGAGTAA CCATGCATCA TCAGGAGTAC GGATAAAATG CTTGATGGTC GGAAGAGGCA 1140 TAAATTCCGT CAGCCAGTTT AGTCTGACCA TCTCATCTGT AACATCATTG GCAACGCTAC 1200 CTTTGCCATG TTTCAGAAAC AACTCTGGCG CATCGGGCTT CCCATACAAT CGATAGATTG 1260 TCGCACCTGA TTGCCCGACA TTATCGCGAG CCCATTTATA CCCATATAAA TCAGCATCCA 1320 TGTTGGAATT TAATCGCGGC CTAGAGCAAG ACGTTTCCCG TTGAATATGG CTCATAACAC 1380 CCCTTGTATT ACTGTTTATG TAAGCAGACA GTTTTATTGT TCATGACCAA AATCCCTTAA 1440 CGTGAGTTTT CGTTCCACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG ATCTTCTTGA 1500

GATCCTTTTT TTCTGCGCGT AATCTGCTGC TTGCAAGA	
GATCCTTTTT TTCTGCGCGT AATCTGCTGC TTGCAAACAA AAAAACCACC GCTACCAGCG	1560
GTGGTTTGTT TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC	1620
AGAGCGCAGA TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG	1680
AACTCIGTAG CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTCCC	
AGTGGCGATA AGTCGTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG	1740
CAGCGGTCGG GCTGAACGGG GGGTTCGTGC ACACAGCCCA GCTTGGAGCG AACGACCTAC	1800
ACCGAACTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC CGAAGGGAGA	1860
AAGGCGGACA GGTATCCGGT AACCCGGTGG	1920
AAGGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT	1980
CCAGGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTCGGGT TTCGCCACCT CTGACTTGAG	2040
CGTCGATTTT TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG	2100
GCCITITAC GGTTCCTGGC CTTTTGCTG CCTTTTGCTC ACATGTTCTT TCCTGCCTTL	
TCCCCTGATT CTGTGGATAA CCGTATTACC GCCTTTGAGT GAGCTGATAC CGCTCGCCGC	2160
AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG CGGAAGAGCG CCTGATGCGG	2220
TATTTTCTCC TTACGCATCT GTGCGGTATT TCACACCGCA TATATGGTGC ACTCTCAGTA	2280
CAATCTGCTC TGATGCCGCA TACTTAAGGTA	2340
CAATCTGCTC TGATGCCGCA TAGTTAAGCC AGTATACACT CCGCTATCGC TACGTGACTG	2400
GGTCATGGCT GCGCCCGAC ACCCGCCAAC ACCCGCTGAC GCGCCCTGAC GGGCTTGTCT	2460
GCTCCCGGCA TCCGCTTACA GACAAGCTGT GACCGTCTCC GGGAGCTGCA TGTGTCAGAG	2520
GITTICACCG TCATCACCGA AACGCGCGAG GCAGCTGCGG TAAAGCTCAT CAGCGTGCTG	2580
GTGAAGCGAT TCACAGATGT CTGCCTGTTC ATCCGCGTCC AGCTCGTTGA GTTTCTCCAG	
AAGCGTTAAT GTCTGGCTTC TGATAAAGCG GGCCATGTTA AGGGCGGTTT TTTCCTGTTT	2640
GGTCACTGAT GCCTCCGTGT AAGGGGGGATT TCTGTTCATG GGGGTAATGA TACCGATGAA	2700
ACGAGAGAG ATGCTCACGA TACCGGTTAG	2760
ACGAGAGAG ATGCTCACGA TACGGGTTAC TGATGATGAA CATGCCCGGT TACTGGAACG	2820
TTGTGAGGGT AAACAACTGG CGGTATGGAT GCGGCGGGAC CAGAGAAAAA TCACTCAGGG	2880
TCAATGCCAG CGCTTCGTTA ATACAGATGT AGGTGTTCCA CAGGGTAGCC AGCAGCATCC	2940
TGCGATGCAG ATCCGGAACA TAATGGTGCA GGGCGCTGAC TTCCGCGTTT CCAGACTTTA	3000
CGAAACACGG AAACCGAAGA CCATTCATGT TGTTGCTCAG GTCGCAGACG TTTTGCAGCA	
GCAGTCGCTT CACGTTCGCT CGCGTATCGG TGATTCATTC TGCTAACCAG TAAGGCAACC	3060
CCGCCAGCCT AGCCGGGTCC TCAACGACAG GAGCACGATC ATGCGCACCC GTGGGGCCGC	3120
CAGCACGATC ATGCGCACCC GTGGGGCCGC	3180

WO 99/42076

CAIGCCGGCG ATAATGGCCT GCTTCTCGCC GAAACGTTTG GTGGCGGGAC CAGTGACGAA	3240
GGCTTGAGCG AGGGCGTGCA AGATTCCGAA TACCGCAAGC GACAGGCCGA TCATCGTCGC	3300
GCTCCAGCGA AAGCGGTCCT CGCCGAAAAT GACCCAGAGC GCTGCCGGCA CCTGTCCTAC	3360
GAGTTGCATG ATAAAGAAGA CAGTCATAAG TGCGGCGACG ATAGTCATGC CCCGCGCCCA	3420
CCGGAAGGAG CTGACTGGGT TGAAGGCTCT CAAGGGCATC GGTCGAGATC CCGGTGCCTA	3480
ATGAGTGAGC TAACTTACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA	3540
CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG GTTTGCGTAT	-
TGGGCGCCAG GGTGGTTTTT CTTTTCACCA GTGAGACGGG CAACAGCTGA TTGCCCTTCA	3600
CCGCCTGGCC CTGAGAGAGT TGCAGCAAGC GGTCCACGCT GGTTTGCCCC AGCAGGCGAA	3660
AATCCTGTTT GATGGTGGTT AACGGCGGGA TATAACATGA GCTGTCTTCG GTATCGTCGT	3720
ATCCCACTAC CGAGATATCC GCACCAACGC GCAGCCCGGA CTCGGTAATG GCGCGCATTG	3780
CECCCAGCGC CATCTGATCG TTGGCAACCA GCATCGCAGT GGGAACGATG CCCTCATTCA	3840
GCATTTGCAT GGTTTGTTGA AAACCGGACA TGGCACTCCA GTCGCCTTCC CGTTCCGCTA	3900
TCGGCTGAAT TTGATTGCGA GTGACATATT TAGGGCTA	3960
TCGGCTGAAT TTGATTGCGA GTGAGATATT TATGCCAGCC AGCCAGACGC AGACGCGCCG	4020
AGACAGAACT TAATGGGCCC GCTAACAGCG CGATTTGCTG GTGACCCAAT GCGACCAGAT GCTCCACGCC CAGTCGCGTA GCGGCCAAACAGCG	4080
GCTCCACGCC CAGTCGCGTA CCGTCTTCAT GGGAGAAAAT AATACTGTTG ATGGGTGTCT	4140
GGTCAGAGAC ATCAAGAAAT AACGCCGGAA CATTAGTGCA GGCAGCTTCC ACAGCAATGG	4200
CATCCTGGTC ATCCAGCGGA TAGTTAATGA TCAGCCCACT GACGCGTTGC GCGAGAAGAT	4260
TGTGCACCGC CGCTTTACAG GCTTCGACGC CGCTTCGTTC TACCATCGAC ACCACCACGC	4320
TGGCACCCAG TTGATCGGCG CGAGATTTAA TCGCCGCGAC AATTTGCGAC GGCGCGTGCA	4380
GGGCCAGACT GGAGGTGGCA ACGCCAATCA GCAACGACTG TTTGCCCCGCC AGTTGTTGTG	4440
CCACGCGGTT GGGAATGTAA TTCAGCTCCG CCATCGCCGC TTCCACTTTT TCCCGCGTTT	4500
TCGCAGAAAC GTGGCTGGCC TGGTTCACCA CGCGGGAAAC GGTCTGATAA GAGACACCGG	4560
CATACTCTGC GACATCGTAT AACGTTACTG GTTTCACATT CACCACCCTG AATTGACTCT	4620
CITCCGGGCG CTATCATGCC ATACCGCGAA AGGTTTTGCG CCATTCGATG GTGTCCGGGA	4680
TCTCGACGCT CTCCCTTATG CGACTCCTGC ATTAGGAAGC ACCCGACTAG TAGGAAGC	4740
CCGTTGAGCA CCGCCGCCGC AAGGAATGGT GCATGCAAGG AGATGGCGCC CAACAGTCCC	4800

CCGGCCACGG GGCCTGCCAC CATACCCACG CCGAAACAAG CGCTCATGAG CCCGAAGTGG	
CGAGCCCGAT CTTCCCCATC GGTGATGTCC GGGATGTGG	4860
CGAGCCCGAT CTTCCCCATC GGTGATGTCG GCGATATAGG CGCCAGCAAC CGCACCTGTG	4920
GCGCCGGTGA TGCCGGCCAC GATGCGTCCG GCGTAGAGGA TCGAGATCTC GATCCCGCGA	4980
AATTAATACG ACTCACTATA GGGGAATTGT GAGCGGATAA CAATTCCCCT CTAGAAATAA	5040
TITTGTTTAA CTTTAAGAAG GAGATATACA TATGGGCCAT CATCATCATC ATCACCTCAT	
CGACATCATC GGGACCAGCC CCACATCCTG GGAACAGGCG GCGGCGGAGG CGGTCCAGCG	5100
GGCGCGGGAT AGCGTCGATG ACATCCGCGT CGCTCGGGTC ATTGAGCAGG ACATGGCCGT	5160
GGACAGCGCC GGCAAGATCA CCTACGGCCCT	5220
GGACAGCGC GGCAAGATCA CCTACCGCAT CAAGCTCGAA GTGTCGTTCA AGATGAGGCC	5280
GGCGCAACCG AGGGGCTCGA AACCACCGAG CGGTTCGCCT GAAACGGGCG CCGGCGCCGG	5340
TACTGTCGCG ACTACCCCCG CGTCGTCGCC GGTGACGTTG GCGGAGACCG GTAGCACGCT	5400
GCICIACCCG CTGTTCAACC TGTGGGGTCC GGCCTTTCAC GAGAGGTATC CGAACGTGAC	
GATCACCGCT CAGGGCACCG GTTCTGGTGC CGGGATCGCG CAGGCCGCCG CCGGGACGGT	5460
CAACATTGGG GCCTCCGACG CCTATCTGTC GGAAGGTGAT ATGGCCGCGC ACAAGGGGCT	5520
GATGAACATC GCGCTAGCCA TCTCCGCTCA GCAGGTCAAC TACAACCTGC CCGGAGTGAG	5580
CGAGCACCTC AAGCTGAACG GAAAACTTGC CCGGAGTGAG	5640
CGAGCACCTC AAGCTGAACG GAAAAGTCCT GGCGGCCATG TACCAGGGCA CCATCAAAAC	5700
CTGGGACGAC CCGCAGATCG CTGCGCTCAA CCCCGGCGTG AACCTGCCCG GCACCGCGGT	5760
AGTICCGCTG CACCGCTCCG ACGGGTCCGG TGACACCTTC TTGTTCACCC AGTACCTGTC	5820
CAMGCAAGAT CCCGAGGGCT GGGGCAAGTC GCCCGGCTTC GGCACCACCG TCGACTTCCC	-
GGCGGTGCCG GGTGCGCTGG GTGAGAACGG CAACGGCGGC ATGGTGACCG GTTGCGCCGA	5880
GACACCGGGC TGCGTGGCCT ATATCGGCAT CAGCTTCCTC GACCAGGCCA GTCAACGGGG	5940
ACTCGGCGAG GCCCAACTAG GCAATAGCTG TGGTAAT	6000
ACTCGGCGAG GCCCAACTAG GCAATAGCTC TGGCAATTTC TTGTTGCCCG ACGCGCAAAG	6060
CATTCAGGCC GCGGCGGCTG GCTTCGCATC GAAAACCCCCG GCGAACCAGG CGATTTCGAT	6120
GATCGACGGG CCCGCCCCGG ACGGCTACCC GATCATCAAC TACGAGTACG CCATCGTCAA	6180
CAACCGGCAA AAGGACGCCG CCACCGCGCA GACCTTGCAG GCATTTCTGC ACTGCGCGAT	6240
CACCGACGGC AACAAGGCCT CGTTCCTCGA_CCAGGTTCAT TTCCAGCCGC TGCCGCCCGC	
GGTGGTGAAG TTGTCTGACG CGTTGATCGC GACGATTTCC AGCGCTGAGA TGAAGACCGA	6300
TGCCGCTACC CTCGCGCAGG AGGCAGGTAA TTTCGAGCGG ATCTCCGGCG ACCTGAAAAC	6360
CCAGATCGAC CAGGTGGAGT CGACGGGAGT TTTT	6420
CCAGATCGAC CAGGTGGAGT CGACGGCAGG TTCGTTGCAG GGCCAGTGGC GCGGCGCGC	6480

GGGGACGGCC GCCCAGGCCG CGGTGGTGCC CTTGGTA TO	
GGGGACGGCC GCCCAGGCCG CGGTGGTGCG CTTCCAAGAA GCAGCCAATA AGCAGAAGCA	6540
GGAACTCGAC GAGATCTCGA CGAATATTCG TCAGGCCGGC GTCCAATACT CGAGGGCCGA	6600
CGAGGAGCAG CAGCAGGCGC TGTCCTCGCA AATGGGCTTT GTGCCCACAA CGGCCGCCTC	
GCCGCCGTCG ACCGCTGCAG CGCCACCCGC ACCGGCGACA CCTGTTGCCC CCCCACCACC	6660
GGCCGCCGCC AACACGCCGA ATCCCCACCACC	6720
GGCCGCCGCC AACACGCCGA ATGCCCAGCC GGGCGATCCC AACGCAGCAC CTCCGCCGGC	6780
CGACCCGAAC GCACCGCCGC CACCTGTCAT TGCCCCAAAC GCACCCCAAC CTGTCCGGAT	6840
CGACAACCCG GTTGGAGGAT TCAGCTTCGC GCTGCCTGCT GGCTGGGTGG AGTCTGACGC	
CGCCCACTTC GACTACGGTT CAGCACTCCT CAGCAAAACC ACCGGGGACC CGCCATTTCC	6900
CGGACAGCCG CCGCCGGTGG CCAATGACAC CCGTATCGTG CTCGGCCGGC TAGACCAAAA	6960
GCTTTACGCC AGCCCCAAA	7020
GCTTTACGCC AGCGCCGAAG CCACCGACTC CAAGGCCGCG GCCCGGTTGG GCTCGGACAT	7080
GGGTGAGTTC TATATGCCCT ACCCGGGCAC CCGGATCAAC CAGGAAACCG TCTCGCTTGA	7140
CGCCAACGGG GTGTCTGGAA GCGCGTCGTA TTACGAAGTC AAGTTCAGCG ATCCGAGTAA	7140
GCCGAACGGC CAGATCTGGA CGGGCGTAAT CGGCTCGCCC GCGGCGAACG CACCGGACGC	7200
CGGGCCCCT CAGCGCTCCT TTGTT	7260
CGGGCCCCCT CAGCGCTGGT TTGTGGTATG GCTCGGGACC GCCAACAACC CGGTGGACAA	7320
GGGCGCGGCC AAGGCGCTGG CCGAATCGAT CCGGCCTTTG GTCGCCCCGC CGCCGGCGCC	7380
GGCACCGGCT CCTGCAGAGC CCGCTCCGGC GCCGGCGCCG GCCGGGGAAG TCGCTCCTAC	
CCCGACGACA CCGACACCGC AGCGGACCTT ACCGGCCTGA GAATTCTGCA GATATCCATC	7440
ACACTGGCGG CCGCTCGAGC ACCACGAGA ACCACGAGA GAATTCTGCA GATATCCATC	7500
ACACTGGCGG CCGCTCGAGC ACCACCACCA CCACCACTGA GATCCGGCTG CTAACAAAGC	7560
CCGAAAGGAA GCTGAGTTGG CTGCTGCCAC CGCTGAGCAA TAACTAGCAT AACCCCTTGG	7620
GGCCTCTAAA CGGGTCTTGA GGGGTTTTTT GCTGAAAGGA GGAACTATAT CCGGAT	2675
	7676

(2) INFORMATION FOR SEQ ID NO:214:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 802 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:

Met Gly His His His His His Val Ile Asp Ile Ile Gly Thr Ser

Pro Thr Ser Trp Glu Gln Ala Ala Glu Ala Val Gln Arg Ala Arg

20		25	30
Asp Ser Val Asp	Asp Ile Arg	Val Ala Aro Va	l Ile Glu Gln Asp Met
			45
Ala Val Asp Ser 50	Ala Gly Lys 55	Ile Thr Tyr Arg	g Ile Lys Leu Glu Val 60
Ser Phe Lys Met 65	Arg Pro Ala (Gln Pro Arg Gly 75	Ser Lys Pro Pro Ser
Gly Ser Pro Glu	Thr Gly Ala 0 85	Gly Ala Gly Thr 90	Val Ala Thr Thr Pro
Ala Ser Ser Pro	Val Thr Leu A	ula Glu Thr Gly 105	Ser Thr Leu Leu Tyr
Pro Leu Phe Asn 1	Geu Trp Gly p	ro Ala Phe His 20	Glu Arg Tyr Pro Asn 125
Val Thr Ile Thr 1	la Gln Gly T 135	hr Gly Ser Gly	Ala Gly Ile Ala Gln
Ala Ala Ala Gly T 145	hr Val Asn II	le Gly Ala Ser 155	Asp Ala Tyr Leu Ser 160
Glu Gly Asp Met A 1	la Ala His Ly 65	s Gly Leu Met	Asn Ile Ala Leu Ala
		170	175 Gly Val Ser Glu His
		203	190
			Tyr Gln Gly Thr Ile 205
		4	Asn Pro Gly Val Asn
		433	er Asp Gly Ser Gly 240
Asp Thr Phe Leu Ph 24		250	255
Trp Gly Lys Ser Pro		203	270
Pro Gly Ala Leu Gly 275	•=		285
Ala Glu Thr Pro Gly 290		3(, o
Gln Ala Ser Gln Arg 305	Gly Leu Gly 310	Glu Ala Gln Le 315	eu Gly Asn Ser Ser

WO 99/42076 PCT/US99/03268

Gly Asn Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala 325 330 335

- Gly Phe Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp 340 345 350
- Gly Pro Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile 355 360 365
- Val Asn Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala 370 375 380
- Phe Leu His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp 385 390 395 400
- Gln Val His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp 405 410 415
- Ala Leu Ile Ala Thr Ile Ser Ser Ala Glu Met Lys Thr Asp Ala Ala
 420
 425
 430
- Thr Leu Ala Glu Ala Gly Asn Phe Glu Arg Ile Ser Gly Asp Leu
 435
 440
 445
- Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala Gly Ser Leu Gln Gly
 450
 455
 460
- Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala Ala Val Val Arg
 470
 475
 480
- Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu Leu Asp Glu Ile Ser 485 490 495
- Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg Ala Asp Glu Glu 500 505 510
- Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe Val Pro Thr Thr Ala 515 520 525
- Ala Ser Pro Pro Ser Thr Ala Ala Ala Pro Pro Ala Pro Ala Thr Pro
 530 535 540
- Val Ala Pro Pro Pro Pro Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro 545 550 555 560
- Gly Asp Pro Asn Ala Ala Pro Pro Pro Ala Asp Pro Asn Ala Pro Pro 565 570 575
- Pro Pro Val Ile Ala Pro Asn Ala Pro Gln Pro Val Arg Ile Asp Asn 580 585 590
- Pro Val Gly Gly Phe Ser Phe Ala Leu Pro Ala Gly Trp Val Glu Ser 595 600 605

200
Asp Ala Ala His Phe Asp Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr 610 615 620
Gly Asp Pro Pro Phe Pro Gly Gln Pro Pro Pro Val Ala Asn Asp Thr 625 630 635 640
Arg Ile Val Leu Gly Arg Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu 645 650 655
Ala Thr Asp Ser Lys Ala Ala Ala Arg Leu Gly Ser Asp Met Gly Glu 660 665 670
Phe Tyr Met Pro Tyr Pro Gly Thr Arg Ile Asn Gln Glu Thr Val Ser 675 680 685
Leu Asp Ala Asn Gly Val Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys 690 695 700
Phe Ser Asp Pro Ser Lys Pro Asn Gly Gln Ile Trp Thr Gly Val Ile 705 710 715 720
Gly Ser Pro Ala Ala Asn Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp 725 730 735
Phe Val Val Trp Leu Gly Thr Ala Asn Asn Pro Val Asp Lys Gly Ala 740 745 750
Ala Lys Ala Leu Ala Glu Ser Ile Arg Pro Leu Val Ala Pro Pro Pro 755 760 765
Ala Pro Ala Pro Ala Pro Ala Glu Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala 770 775 780
Gly Glu Val Ala Pro Thr Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu 785 790 795 800
Pro Ala
INFORMATION FOR SEQ ID NO:215:
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 454 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single

- (2) INF
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Genomic DNA
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:215:

GCGGTCCGCG	CGGTGACGAA	GCTGACCCCG	TCGCAGATCA	TCAGGAAGCC	CTGATCGCGG GTCCAGCAGC GGCGATGGCG CACCAGGACC GATGGCCTCC	120

CCGATCGCGA TCAGCTGCTT ACCGACCGGC GGGTGAACCA CCAGGCCGTA CCCGGGGTTG TCTTCCACCC CATGGTTGTT CAGCACCTGC CAGGCCTGGC GGTGCGTAAT GCTTCTCGTC GAAGATGGGG GTGCCGGCAT CCGTCACCGA GCCC (2) INFORMATION FOR SEQ ID NO:216: (i) SEQUENCE CHARACTERISTICS:	360 420 454
(A) LENGTH: 470 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: Genomic DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:	
TGCAGAAGTA CGGCGGATCC TCGGTGGCCG ACGCCGAACG GATTCGCCGC GTCGCCGAAC GCATCGTCGC CACCAAGAAG CAAGGCAATG ACGTCGTCGT CGTCGTCTT GCCATGGGGG ATACCACCGA CGACCTGCTG GATCTGGCTC AGCAGGTGTG CCCGGCGCCG CCGCCTCGGG GCCGGTCGAAC GCCGGTCGT TCACCGGTTC GCAGGCCGA GCCATGGCCA CCGGCACCCA CGGCAACGCC AAGATCATCG ACGTCACGCC GGGGCGGCTG CAAACCGCCC TTGAGGAAGG GCGGGTCGTC TTGGTGGCCG GATTCCAAGG GGTCAGCCAC GACACCAACGC ATGTCACGAC GTTGGGCCG GGCGGCTCG ACACCCACCGC CGTCGCCATG	60 120 180 240 300 360 420
(2) INFORMATION FOR SEQ ID NO:217:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 279 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: Genomic DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:	
GGCCGGCGTA CCCGGCCGGG ACAAACAACG ATCGATTGAT ATCGATGAGA GACGGAGGAA TCGTGGCCCT TCCCCAGTTG ACCGACGAGC AGCGCGCGGC CGCGTTGGAG AAGGCTGCTG CCGCACGTCG AGCGCGAGCA GAGCTCAAGG ATCGGCTCAA GCGTGGCGGC ACCAACCTCA CCCAGGTCCT CAAGGACGCG GAGAGCGATG AAGTCTTGGG CAAAATGAAG GTGTCTGCGC TGCTTGAGGC CTTGCCAAAG GTGGGCAAGG TCCAGGCGC	60 120 180 240
(2) INFORMATION FOR SEQ ID NO:218:	279
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 219 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: Genomic DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:218:	

ACACGGTCGA ACTCGACGAG CCCCTCGTGG AGGTGTCGAC CGACAAGGTC GACACCGAAA	60
TCCCTCGCCG GCCGCGGTG TGCTGACCAA GATCATCGCC CAAGAAGATG ACACGGTCGA	120
GGTCGGCGGC GAGCTCTCTG TCATTGGCGA CGCCCATGAT GCCGGCGAGG CCGCGGTCCC	180
GGCACCCCAG AAAGTCTCTG CCGGCCCAAC CCGAATCCA	219
(2) INFORMATION FOR SEQ ID NO:219:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 342 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: Genomic DNA	
(11) MODECOME TIPE: GENOMIC DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:219:	
TCGCTGCCGA CATCGGCGCC GCGCCCGCCC CCAAGCCCGC ACCCAAGCCC GTCCCCGAGC	60
CAGCGCCGAC GCCGAAGGCC GAACCCGCAC CATCGCCGCC GGCGGCCCAG CCAGCCGGTG	120
CGGCCGAGGG CGCACCGTAC GTGACGCCGC TGGTGCGAAA GCTGGCGTCG GAAAACAACA	180
TCGACCTCGC CGGGGTGACC GGCACCGGAG TGGGTGGTCG CATCCGCAAA CAGGATGTGC	240
TGGCCGCGGC TGAACAAAAG AAGCGGGCGA AAGCACCGGC GCCGGCCG	300
CCGCGCCGGC CCCGAAAGCG CCGCCTGAAG ATCCGATGCC GC	342
(2) INFORMATION FOR SEQ ID NO:220:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 515 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: Genomic DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:	
GGGTCTTGGT CAGTATCAGC GCCGACGAGG ACGCCACGGT GCCCGTCGGC GGCGAGTTGG	60
CCCGGATCGG TGTCGCTGCC GACATCGGCG CCGCGCCCGC CCCCAAGCCC GCACCCAAGC	120
CCGTCCCCGA GCCAGCGCCG ACGCCGAAGG CCGAACCCGC ACCATCGCCG CCGGCGGCCC	180
AGCCAGCCGG TGCGGCCGAG GGCGCACCGT ACGTGACGCC GCTGGTGCGA AAGCTGGCGT	240
CGGAAAACAA CATCGACCTC GCCGGGGTGA CCGGCACCGG AGTGGGTGGT CGCATCCGCA	300
AACAGGATGT GCTGGCCGCG GCTGAACAAA AGAAGCGGGC GAAAGCACCG GCGCCCTGAG	360
CGCTTCATCA CCCGGTTAAC CAGCTTGCCC CAGAAGCCGG CTTCGACCTC TTCGCGGGTC	420
TTGGTCCGCT GCAGGCGGTC GGCGAGCCAG TTCAGGTTAG GCGGCCGAAA TCTTCCAGTT CGCCAGGAAG GGCACCCGGA ACAGGGTCCG CACCC	480
SOCIAL GOLACCEGGA ACAGGGTCCG CACCC	515
(2) INFORMATION FOR SEQ ID NO:221:	
10.221,	
(i) SEQUENCE CHARACTERISTICS:	

(A) LENGTH: 557 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:

GACAGCTGGC ACCAGGCCAC CGTTCGCAGC ACCAGGCGCT TGCAGCAGAA AGGAGCAGGT	GGTGCAGATT GGCGCGCAA GGACATCGAA CGCCGCCGGA CCAGCTGGTG TAGCGCCGCA GATCGCCGAG CAGCGCATCG CGAGGTGCGC CGAGAGT	AAGCTTCAGG GACGCTGCCA ACCGCCGAGC GCTCAGGCCA CGAACCAAGC TTGCGGTCGA	ACCAGCGTCA TCAATGTGCG AGGCCACCGA AGAGCGTCGA AGAAGGCCGT TGCTCAGCCA TGAGTGAGCT	ATTGGAGATG CCAAGCCCTG ATACAACAAC AGACCTCAAG CGAACGAAAT GCTCGAGCAC	CGACTCAACC ACGCTGGCCG GCCGCCGAGG ACGCTGCATG GCGATGGTGC GCGAAGATGC	60 120 180 240 300 360 420 480 540
CIGAACTIGC	CGAGAGT					557

(2) INFORMATION FOR SEQ ID NO:222:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 223 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:

CAGGATAGGT	TTCGACATCC	ACCTGGGTTC	CGCACCCGGT	CCCCC3 CCCm	GTGATAGGCC	
AGAGGTGGAC	CECCCCCCCC	003.003.003		GCGCGACCGI	GTGATAGGCC	60
T0010010	CIGCOCCOAC	CGACGATCGA	TCGAGGAGTC	AACAGAAATG	GCCTTCTCCG	120
TCCAGATGCC	GGCACTCGGT	GAGAGCGTCA	CCGACCCCAC	COMM's COCCO	TGGCTCAAAC	120
AGGAAGGCGA	CACCCCCCAA	CTCCCCCCCC	CCONGGGAC	GGTTACCCGC	TGGCTCAAAC	180
	CACGGTCGAA	CTCGACGAGC	CCCTCGTGGA	GGT		223
						443

- (2) INFORMATION FOR SEQ ID NO:223:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 578 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:

(2) INFORMATION FOR SEQ ID NO:224:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 484 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

TCAAGGATCG GCGATGAAGT GCAAGGTCAA TCGTGGCCTC	GCTCAAGCGT CTTGGGCAAA GGCGCAGGAG GGTGACCGTC ACGATGCGGG	GGCGGCACCA ATGAAGGTGT ATCATGACCG AGCGCAAGGC	ACCTCACCCA CTGCGCTGCT AGCTGGAAAT CCTGCTGGAA	ACGTCGAGCG GGTCCTCAAG TGAGGCCTTG TGCGCCCCAC AAGTTCGGCT	CAGTTGACCG CGAGCAGAGC GACGCGGAGA CCAAAGGTGG CCCGCCGCCT CCGCCTAACC ATACGGGGGA GGGGGAACCC	60 120 180 240 300 360 420 480
--	--	--	--	--	---	---

- (2) INFORMATION FOR SEQ ID NO:225:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 537 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

CAAGCCAATC CCAAGCAGCA CCCCGGC 537	ACGAAGTACA CGCCGGACCA CTTGGTGGCG CAAGCCAATC CCAAGCAGCA CCCCGGC 537	ACTTGGTACT GACGTCGCGC CGCCAGCCA CGCCGCACTG TCCCCAGCCA CGGTCTTTGC 4	GCGGTCTCGG ACGCGAACCA CGGAGCGTAG GTCGGGAGAT CCGATCGATC 30	TCGGGCAGCG GAACAACCGA GTTGCGAGCG AGGGTGTAGT GCCAGAGCGA GCGCACGGCG 240	GTCAGGCCCG CGTGGAAGTG GAACCCCTTTC 180	AGGACTGGCA CCGCCAGCCA CCACATGGGG	AACGACATGG GTCAGGCCCG TCGGGCAGCG GCGGTCTCGG CCCAGCGCAT ACTTGGTACT	GCCAGGTCCA CGTGGAAGTG GAACAACCGA ACGCGAACCA ACCCGCTGGG GACGTCGCGC	CGGTTTGGAT GAACGCTTTG GTTGCGACCG CGGAGCGTAG AAGCACGTCA	TCCCAAGGGT GCGGTGTAGT ACCGCTTGAC GTGGCCAGAT CGCCGCACTG	TGGCGTACAG GGTAGTTGCC GCCAGAGCGA CGACCGCATG AGACCGCGAA TCCCCAGCCA	CACCGGCCGC TGCGGAATTC GCGCACCGCG CCGATCGATC CGGGATCAAC CGGTCTTTGC	
GTCAGGCCCG CGTGGAAGTG GAACGCTTTG GCGGTTTTGCC TGCGGAATTC 180 TCGGGCAGCG GAACAACCGA GTTGCGACCG ACCGCTTGAC CGACCGCATG CCGATCGATC 300 GCGGTCTCGG ACGCGAACCA CGGAGCGTAG GTGGCCAGAT AGACCGCGAA CGGGATCAAC 360 ACTTGGTACT GACGTCGCC CGCCACGTCG CGCCACGCCA	GTCAGGCCCG CGTGGAAGTG GAACGCTTTG GCGGTGTTAGT GCCAAGGGT GGTAGTTGCC TGCGGAATTC TCGGGCAGCG GAACAACCGA GTTGCGACCG ACCGCTTGAC CGACCGCATG CCGATCGATC GCGGTCTCGG ACGCGAACCA CGGAGCGTAG GTGGCCAGAT AGACCGCGAA CGGGATCAAC CCCAGCGCAT ACCCGCTGGG AAGCACCACAG GTGGCCAGAT AGACCGCGAA CGGGATCAAC 360	GTCAGGCCCG CGTGGAAGTG GAACGCTTTG GCGGTGTAGT GCCAGAGCGA GCGCACGGCG TCGGGCAGCG GAACAACCGA GTTGCGACCG ACCGCTTGAC CGACCGCATG CCGATCGATC 3	GTCAGGCCCG CGTGGAAGTG GAACGCTTTG GCGGTGTAGT GCCAGAGCGA GCGCACGGCG 24	GTCAGGCCCG CGTGGAAGTG GAACCCCTTTC 180		AGGACTGGCA CCGCCAGCCA CCACATCGCG GGCGTGCCGA CCAGCATCTC GGCCTTGACG CACGACTGTG CGCCGCAGCC TGCAACGTCT TGCTGACGA CCAGCATCTC GGCCTTGACG	AACGACATGG	GCCAGGTCCA	CGCTTTTCCA	TGCTGGTCGA	TGGCGTACAG	CACCGGCCGC	
AACGACATGG GCCAGGTCCA CGGTTTGGAT TCCCAAGGGT GGTAGTGCC TGCGGAATTC 180 GTCAGGCCCG CGTGGAAGTG GAACGCTTTG GCGGTTTGAC GCCAGAGCGA GCGCACGGCG 240 GCGGTCTCGG ACGCGAACCA CGGAGCGTAG GTGGCCAGAT ACCCGCTGGG AAGCACGTCA CGGCGCATG CCGATCGATC CCCAGCGCAT ACCCGCTGGG AAGCACGTCA CGCCACGCAT TCCCCAGCCA CGGTCTTTGC 420 ACTTGGTACT GACGTCGCC CGCCACGTGA CGCCACGCCA	AACGACATGG GCCAGGTCCA CGGTTTGGAT TCCCAAGGGT GGTAGTTGCC TGCGGAATTC 180 GTCAGGCCCG CGTGGAAGTG GAACGCTTTG GCGGTTTAGT GCCAGAGCGA GCGCACGGCG 240 GCGGTCTCGG ACGCGAACCA CGGAGCGTAG GTGCCAGAT ACCCGCTTGAC CGACCGCAT CCGATCGATC 300 CCCCAGCGCAT ACCCGCTGGG AAGCACCTCA GTGCCAGAT AGACCGCGAA CGGGATCAAC 360	AACGACATGG GCCAGGTCCA CGGTTTGGAT TCCCAAGGGT GGTAGTTGCC TGCGGAATTC GTCAGGCCCG CGTGGAAGTG GAACGCTTTG GCGGTTTAGT GCCAGAGCGA GCGCACGGCG TCGGGCAGCG GAACAACCGA GTTGCGACCG ACCGCTTGAC CGACCGCATG CCGATCGATC GCGGTCTCGG ACGCGAACCA CGGAGCGTAG CCGATCGATC 3	AACGACATGG GCCAGGTCCA CGGTTTGGAT TCCCAAGGGT GGTAGTTGCC TGCGGAATTC 18 GTCAGGCCCG CGTGGAAGTG GAACGCTTTG GCGGTGTAGT GCCAGAGCGA GCGCACGGCG 24	AACGACATGG GCCAGGTCCA CGGTTTGGAT TCCCAAGGGT GGTAGTTGCC TGCGGAATTC 180	AACGACATGG GCCAGGTCA CGCTTTGGAR TGCTGGTCGA TGGCGTACAG CACCGGCCGC 120	TOTAL COURT COUNTY CAROLOG	CACGACTGTG	CGCCGCAGCCA	CCACATCGCG	GGCGTGCCGA	CCAGCATCTC	GGCCTTGACG	60

- (2) INFORMATION FOR SEQ ID NO:226:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 135 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single_
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:

(2) INFORMATION FOR SEQ ID NO:227:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 156 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

Gln Lys Tyr Gly Gly Ser Ser Val Ala Asp Ala Glu Arg Ile Arg Arg Val Ala Glu Arg Ile Val Ala Thr Lys Lys Gln Gly Asn Asp Val Val 25 Val Val Val Ser Ala Met Gly Asp Thr Thr Asp Asp Leu Leu Asp Leu 40 Ala Gln Gln Val Cys Pro Ala Pro Pro Pro Arg Glu Leu Asp Met Leu Leu Thr Ala Gly Glu Arg Ile Ser Asn Ala Leu Val Ala Met Ala Ile Glu Ser Leu Gly Ala His Ala Arg Ser Phe Thr Gly Ser Gln Ala Gly 90 Val Ile Thr Thr Gly Thr His Gly Asn Ala Lys Ile Ile Asp Val Thr 105 Pro Gly Arg Leu Gln Thr Ala Leu Glu Glu Gly Arg Val Val Leu Val 120 Ala Gly Phe Gln Gly Val Ser Gln Asp Thr Lys Asp Val Thr Thr Leu 125 135 Gly Arg Gly Gly Ser Asp Thr Thr Ala Val Ala Met 150

- (2) INFORMATION FOR SEQ ID NO:228:
- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 92 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

 Pro
 Ala
 Tyr
 Pro
 Ala
 Gly
 Thr
 Asn
 Asn
 Asp
 Arg
 Leu
 Ile
 Ser
 Met
 Arg

 Asp
 Gly
 Gly
 Ile
 Val
 Ala
 Leu
 Pro
 Gln
 Leu
 Thr
 Asp
 Glu
 Gln
 Arg
 Ala

 Ala
 Ala
 Leu
 Glu
 Ala
 Ala
 Ala
 Ala
 Arg
 Arg
 Ala
 Arg
 Arg
 Ala
 Arg
 Arg
 Ala
 Arg
 Arg
 Arg
 Ala
 Arg
 Arg
 Arg
 Ala
 Arg
 A

- (2) INFORMATION FOR SEQ ID NO:229:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 72 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

- (2) INFORMATION FOR SEQ ID NO:230:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 113 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

(2) INFORMATION FOR SEQ ID NO:231:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 118 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:231:

 Val
 Leu
 Val
 Ser
 Ile
 Ser
 Ala
 Asp
 Glu
 Asp
 Ala
 Thr
 Val
 Pro
 Val
 Gly
 Glu
 Inc
 I

(2) INFORMATION FOR SEQ ID NO:232:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 185 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:232:

Asp Pro Lys Val Gln Ile Gln Gln Ala Ile Glu Glu Ala Gln Arg Thr 10 His Gln Ala Leu Thr Gln Gln Ala Ala Gln Val Ile Gly Asn Gln Arg Gln Leu Glu Met Arg Leu Asn Arg Gln Leu Ala Asp Ile Glu Lys Leu 40 Gln Val Asn Val Arg Gln Ala Leu Thr Leu Ala Asp Gln Ala Thr Ala 55 Ala Gly Asp Ala Ala Lys Ala Thr Glu Tyr Asn Asn Ala Ala Glu Ala 70 Phe Ala Ala Gln Leu Val Thr Ala Glu Gln Ser Val Glu Asp Leu Lys 85 Thr Leu His Asp Gln Ala Leu Ser Ala Ala Ala Gln Ala Lys Lys Ala 105 Val Glu Arg Asn Ala Met Val Leu Gln Gln Lys Ile Ala Glu Arg Thr 120 Lys Leu Leu Ser Gln Leu Glu Gln Ala Lys Met Gln Glu Gln Val Ser 135 Ala Ser Leu Arg Ser Met Ser Glu Leu Ala Ala Pro Gly Asn Thr Pro 150 155 Ser Leu Asp Glu Val Arg Asp Lys Ile Glu Arg Arg Tyr Ala Asn Ala 165 Ile Gly Ser Ala Glu Leu Ala Glu Ser 180

- (2) INFORMATION FOR SEQ ID NO:233:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 71 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:233:

 Val
 Ser
 Thr
 Ser
 Trp
 Val
 Pro
 His
 Pro
 Val
 Arg
 Arg
 Val
 1le

 1
 Image: Ser
 Trp
 Thr
 Ser
 Image: Ser

(2) INFORMATION FOR SEQ ID NO:234:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 182 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

Glu Val His Leu Pro Val Asp Val Gly Glu Pro Arg Gln Pro Thr Gly 10 Ala Val Ala Asp Gln Asp His Arg Ile Thr Pro Val Pro Ala His Lys 20 25 His Thr Pro Pro Arg Val Cys Gln Asp Trp His Arg Gln Pro Pro His 40 Arg Gly Arg Ala Asp Gln His Leu Gly Leu Asp Ala Arg Leu Cys Ala Ala Ala Cys Asn Val Leu Leu Val Asp Gly Val Gln His Arg Pro Gln 70 75 Arg His Gly Pro Gly Pro Arg Phe Gly Phe Pro Arg Val Val Ala 90 Cys Gly Ile Arg Gln Ala Arg Val Glu Val Glu Arg Phe Gly Gly Val 105 Leu Pro Glu Arg Ala His Gly Val Gly Gln Arg Asn Asn Arg Val Ala 120 Thr Asp Arg Leu Thr Asp Arg Met Pro Ile Asp Arg Gly Leu Gly Arg 135 Glu Pro Arg Ser Val Gly Gly Gln Ile Asp Arg Glu Arg Asp Gln Pro 150 155 Gln Arg Ile Pro Ala Gly Lys His Val Thr Pro His Cys Ser Gln Pro 165 170 Arg Ser Leu His Leu Val 180

- (2) INFORMATION FOR SEQ ID NO:235:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 160 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:

(2) INFORMATION FOR SEQ ID NO:236:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 178 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:236:

Asp Trp His Arg Gln Pro Pro His Arg Gly Arg Ala Asp Gln His Leu Gly Leu Asp Ala Arg Leu Cys Ala Ala Cys Asn Val Leu Leu Val 25 Asp Gly Val Gln His Arg Pro Gln Arg His Gly Pro Gly Pro Arg Phe 40 Gly Phe Pro Arg Val Val Ala Cys Gly Ile Arg Gln Ala Arg Val 55 Glu Val Glu Arg Phe Gly Gly Val Val Pro Glu Arg Ala His Gly Val 70 Gly Gln Arg Asn Asn Arg Val Ala Thr Asp Arg Leu Thr Asp Arg Met Pro Ile Asp Arg Gly Leu Gly Arg Glu Pro Arg Ser Val Gly Gln Ile Asp Arg Glu Arg Asp Gln Pro Gln Arg Ile Pro Ala Gly Lys His 120 Val Thr Pro His Cys Pro Gln Pro Arg Ser Leu His Leu Val Leu Thr Ser Arg Arg His Val Glu Arg Gln Arg His Arg Ala Glu Gln His 150 Glu Val His Ala Gly Pro Leu Gly Gly Ala Ser Gln Ser Gln Ala Ala 155 170 Pro Arg

(2) INFORMATION FOR SEQ ID NO:237:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 271 base pairs
 - (B) TYPE: nucleic acid

(C)	STRANDEDNESS:	single
	TODOLOGIC 1:	•

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

ATGCCAAGCC GC GTTGGCTGAA GA CCGACAAGGT GC CCGACGAGGA CC	AAGATCGGG GACACCGAG GCCACGGTG	GATTCGGTTC ATCCCGTCCC CCCGTCGGCG	AGGTTGACGA CGGTGGCTGG GCGAGTTGGC	GCCACTCGTG GGTCTTGGTC	GAGGTGTCCA	60 120 180 240
AGATCGGCGC CC		CCCATCGGCG	GCGAGIIGGC	CCGGATCGGT	GTCGCTGCCG	240
		CCCAAGCCCC	C			271

- (2) INFORMATION FOR SEQ ID NO:238:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 89 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

- (2) INFORMATION FOR SEQ ID NO:239:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 107 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:

GAGGTAGCGG ATGGCCGGAG GAGCACCCCA GGACCGCGC CGAACCGCGG GTGCCGGTCA 60 TCGATATGTG GGCACCGTTC GTTCCGTCCG CCGAGGTCAT TGACGAT 107

(2) INFORMATION FOR SEQ ID NO:240:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 339 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

GGCTATTGCC TACCCCGACG	CGGGTGGCCG GCTCGTTTTG GTGTCAGCGG	ATGGGGTTTT GCACCAGTGG	GGCGACTTGG ATGCAAACGT	ATCAGCCGGA CCGTGTGCGA	GGCGCTTGTG CATGACGAAA CGGCGAGAAG CCCACAGTTT GGGTGGTTGC	120 180
			COLCUIGA			330

- (2) INFORMATION FOR SEQ ID NO:241:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 112 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

 Met
 Lys
 Leu
 Lys
 Phe
 Ala
 Arg
 Leu
 Ser
 Thr
 Ala
 Ile
 Leu
 Gly
 Cys
 Ala

 Ala
 Ala
 Leu
 Val
 Phe
 Pro
 Ala
 Ser
 Val
 Ala
 Ser
 Ala
 Asp
 Pro
 Pro
 Asp

 Pro
 His
 Gln
 Pro
 Asp
 Met
 Thr
 Lys
 Gly
 Tyr
 Cys
 Pro
 Gly
 Arg
 Trp

 Gly
 Phe
 Gly
 Asp
 Leu
 Ala
 Val
 Cys
 Asp
 Gly
 Fro
 Gly
 Fro
 Asp
 Gly
 Trp

 Gly
 Phe
 Gly
 Asp
 Leu
 Ala
 Val
 Cys
 Asp
 Gly
 Gly
 Trp
 Asp
 Gly
 Trp
 Asp
 Gly
 Gly
 Trp
 Asp
 Gly
 Fro
 Asp
 Gly
 Fro
 Asp
 Gly
 Fro
 Asp
 Gly
 Fro
 Fro
 Fro
 Fro
 Fro
 Fro
 Fro
 Fro
 Fro
 F

- (2) INFORMATION FOR SEQ ID NO:242:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 371 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

180

240 300 317

GTGACCACGG TGGGCCTGCC ACCAACCCGG GCAGCGGCAG CCGCGGCGGC GCCGGCGGCT CCGGCGGCAA CGGTGGCGCC GGGGGTAACG CCACCGGCTC AGGCGGCAAC GGCGGCAACG GCGGCAACG GCGGCAACG GCGGCAACG GCGGCAACG GCGGCAACG GCGGCAACG GCGGCAACG GCGGCAACG GCGCGCAACG GCGCCAACGGC CGGCAACGG GCGCCAACGG CGCAACGGC CGGCAACGG CGCGCAACGG GCGCGCAACG GCGCGCAACG GCGCGCCGC TGGCGCTGC GGCGCCGCG GCGCCCGCC	120 180 240
(2) INFORMATION FOR SEQ ID NO:243:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 424 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA	
(vi) CECTURED	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:	
GTCCGGGTCC CACCACCGCG CCGGCGCCC CCTAGCGGCC GGGCGCACCA GCCCCTTTTC TTGACTCGTT CAAGAAAAGG GCCTTCTGTT TGGTCGGCCA TGTTGGCATG ATCGTGACCC ATGGGCAACA TCGACGTCGA CATCTCGGCC AAGGTCTAGC TCCATGCGAA TCGCCGCCGC GCGTCGGAG CCCGGGGTTG TGTCCTACGC GGTGCTCGGA AAGGGTCGG TCGGCAACAT CGGCGCGCC CCAATGGGGT GGGAGGCGGT GTTCACCAAG CCGTTCCAGG CGTTTTGGGT CGACC CCGACGTCAACA ACTGGGTGGA CATCGGGCTG CCCGAGGTGT ACGACGATCC	60 120 180 240 300 360 420
(2) INFORMATION DOLLAR	424
(2) INFORMATION FOR SEQ ID NO:244: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 317 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:	
GCGATGGCGG CCGCGGGTAC CACCGCCAAT GTGGAACGGT TTCCCAACCC CAACGATCCT TTGCATCTGG CGTCAATTGA CTTCAGCCCG GCCGATTTCG TCACCGAGGG CCACCGTCTA AGGGCGGATG CGATCCTACT GCGCCGTACC CACCGCTCTA	60
AGGGCGGATG CGATCCTACT GCGCCGTACC GACCGCTGC CTTTCCCGGA CCACCGTCTA	120
TGGGACTTCC TCC3 CTCC3 CTCC3	100

(2) INFORMATION FOR SEQ ID NO:245:

AGGGCGGATG CGATCCTACT GCGCCGTACC GACCGGCTGC CTTTCGCCGA GCCGCCGGAT

TGGGACTTGG TGGAGTCGCA GTTGCGCACG ACCGTCACCG CCGACACGGT GCGCATCGAC

GTCATCGCCG ACGATATGCG TCCCGAACTG GCGGCGGCGT CCAAACTCAC CGAATCGCTG

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 422 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

TGGCGTATGC	GCTTCGCAGC	CGGTGCCGCG	TCAACGCGCC	GGAGGCAATC	GCTTCGCTGC	60
CGAGGAATGG	TTCGATCACG	ATCGCAGTGT	GCCGTCGTGC	ACCGACACCG	CCGTCCAACG	120
TGAACTGAGG	GCGGAAAATC	GGCCGAAATC	TCGCCCTCAG	TTCACGCTCG	GCGCCTAACG	180
GTTCTGGAAG	TTGGGTGCGC	GCTTCTCGGC	GAACGCGCGC	GGGCCTTCCT	TGGCGTCGTC	240
GGACAGGAAG	ACCTTGATGC	CGATCTGGGT	GTCGATCTTG	AACGCCTCGT	TTTCGGGCAT	300
GCACTCGGTC	TCGCGGATGG	ACCGCAAGAT	GGCCTGCACG	GCCAGGGGTC	CGTTAGCCGA	360
GATGGCGTCG	GCAAGTTCTA	GAACCTTGGT	CAACGCCTGG	CCGTCGGGCA	CACGTGGCCG	420
AT		_				422
						7444

- (2) INFORMATION FOR SEQ ID NO:246:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 426 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:246:

GCGTGCCGCT	GAACACCAGC	CCGCGGCTGC	CAGATCTCCC	GGACTCGGTA	GTGCCGCCGG	60
TGGCGTCGTT	GCTCTCCTGA	CGGGGCGCGG	CGACCATAAG	GTCGCTAATG	CCCAGGTAGC	120
GGCCCAGGTG	CATGGAGTCG	ATGATGATGC	GACTCTCCAG	CTCGCCGACC	GGGAGCTTGG	180
CATCGGGCCT	GATCAGCCAG	GACGCGTAGG	ACAAGTCGAT	CGAATGCATA	GTGGCCTCCA	240
GAGTGGCCGT	GCCACTTCCG	GCGTGCTCCA	CGGCAAATGC	CTTGATTTCT	AGCTCCGCGT	300
AGTGTTCCCG	CATCGCCTGC	GGGATGAATG	GGAACCGCAG	GATGGCGACA	AACGGGTCTG	360
ACCTCAGGTT	TGCCGCTTTG	CGCACAGTGG	TCGACAGCCG	GTACTCGGCA	TAAATGCTGG	420
CCCCGA						426

- (2) INFORMATION FOR SEQ ID NO:247:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 327 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

AGACCGGCGA GGGTGTGGTC	GCTGCCGCG	GCATTGTCGA	TAATCTGCGC	TGGGTCGACG	60
CGCCGATCAA CTAGTGAGGC	GCAACGCTAG	GCTTTGGGAT	ACCCACAGCT	AAAAAGTTTA	120
TCAAAGAAAC GAAGAAGGTT	GCCATGAGCA	CTGTTGCCGC	CTACGCCGCC	ATGTCGGCGA	180
CCGAACCCCT GACCAAGACC	ACGATCACCC	GTCGCGACCC	GGGCCCGCAC	GACATGGCGA	240
TCGACATCAA ATTCGCCGGA	ATCTGTCGCT	CGGACATCCA	TACCGTCCAA	ACCGAATGGG	300
GGCAACCGAA TTTACCTGTG	GTCCCTG				327

(2) INFORMATION FOR SEQ ID NO:248:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 123 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

- (2) INFORMATION FOR SEQ ID NO:249:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 104 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

 Met
 Ala
 Ala
 Gly
 Thr
 Thr
 Ala
 Asn
 Val
 Glu
 Arg
 Phe
 Asn
 Pro
 Asn
 Pro
 Asn
 Fro
 Asn
 Asn
 Asp
 Phe
 Asp
 A

- (2) INFORMATION FOR SEQ ID NO:250:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 41 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

- (2) INFORMATION FOR SEQ ID NO:251:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:

Val Pro Leu Asn Thr Ser Pro Arg Leu Pro Asp Leu Pro Asp Ser Val

Solution 10 15

Val Pro Pro Val Ala Ser Leu Leu Ser

20 25

- (2) INFORMATION FOR SEQ ID NO:252:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 61 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:

Gln Thr Glu Trp Gly Gln Pro Asn Leu Pro Val Val Pro	
50 55 60	
(2) TYPOTHER TON THE COLUMN TO THE COLUMN TO THE COLUMN	
(2) INFORMATION FOR SEQ ID NO:253:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 213 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(a) 10102001. Itheat	
(ii) MOLECULE TYPE: CDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:	
GCTTGGAGCC CTGGAGCGAC GGTGTGGGTC TGGGGGTCGA TTCGTTCTCG GCGAAAGTCA	50
ACTAAAGACC ACGTTGACAC CCAACCGGCG GCCCGGCATG GGCCGTCGCG GCGTAGAAGC	60 120
TTTGACCGCG GCGCGAAACG TTCGCTGCTG CGGCCCATGC AGATCGCACA CGCTTGCTTG	180
AACATCGGGT GGAGCCGGTG GTAACGCCAG GCT	213
(2) INFORMATION FOR SEQ ID NO:254:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 367 base pairs (B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:	
CCGAGCTGCT GTTCGGCGCC GGCGGTGCGG GCGGCGCGGG TGGGGCGGGC	
GGCCCGGTGC TACCGGCGGG ACCGGCGGAC ACGGCGGAGT CGGCGGCGAC GGCGGATGGC	60 120
TGGCACCCGG CGGGGCCGGC GGGGCCGGCG GGCAGGCGG GGCAGGTGGT GCCCGCAGCG	180
ATGGTGGCGC GTTGGGTGGT ACCGGCGGGA CGGGCGGTAC CGGCGGCGCC GGTGGCGCCG	240
GCGGTCGCGG CACACTGCTG CTGGGCGCTG GCGGACAGGG CGGCCTCGGC GGCGCCGGCG	300
GACAAGGCGG CACCGGCGGG GGCCGGCGGA GATGGCGTTC TGGGGGGTGT CAGTGGCACT	360
GGIGGIA	367
(2) INFORMATION FOR SEQ ID NO:255:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 420 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:255:	
AAGGCGTGAT TGGCAAGGCG ACCGCGCAGC GGCCCGTAGC CGCGGGACGG CCCAGGCCCC	60

GACCGCAGCG GCCGGTGTCT GACCGGGTCA GCGACCAGCG GCGCTGACCG TGCCGCTCGT 120 CTACTTCGAC GCCAGCGCCT TCGTCAAACT TCTCACCACC GAGACAGGGA GCTCGCTGGC 180

GTCCGCTCTA	TGGGACGGCT	GCGACGCCGC	ATTETCENC	CCCCTCCCC	ACCCCGAAGT	_
CCCCCCCCC	00000000		ATTOIC CAME	COCCIOGCCI	ACCCCGAAGT	240
CCGCGCCGCA	CICGCIGCAA	CGGGCCGCAA	TCACGACCTA	ACCGAATCCG	AGCTCGCCGA	300
CGCCGAGCGT	GACTGGGAGG	ACTITICATOR	CCCNCCCCCC	C1 CBCC1 1 CT	CACCGCGACG	300
		1101101000	COCMCCCGCC	CAGTCGAACT	CACCGCGACG	360
GITGAACAGC	ACGCCGGGCA	CCTCGCCCGA	ACACATGCCT	TACGCGGAGC	CGACACCGTT	400
					conscience (Q1)	420

- (2) INFORMATION FOR SEQ ID NO:256:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 299 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:

- (2) INFORMATION FOR SEQ ID NO:257:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:

Leu Glu Pro Trp Ser Asp Gly Val Gly Leu Gly Val Asp Ser Phe Ser 1 5 10 15
Ala Lys Val Asn 20

- (2) INFORMATION FOR SEQ ID NO:258:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 121 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:

Glu Leu Leu Phe Gly Ala Gly Gly Ala Gly Gly Ala Gly Gly Ala Gly 1 5 10 15
Thr Asp Gly Gly Pro Gly Ala Thr Gly Gly Thr Gly Gly His Gly Gly

- (2) INFORMATION FOR SEQ ID NO:259:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 34 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:

- (2) INFORMATION FOR SEQ ID NO:260:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 99 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:260:

Gly Gly Ala Gly Gly Phe Gly Asp His Gly Thr Leu Gly Thr Gly Gly 85 90 95 Ala Gly Gly

- (2) INFORMATION FOR SEQ ID NO:261:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 282 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:

TCCTGTTCGG	CGCCGGCGGG	GTGGGCGGTG	TTCCCCCTCA	CCCTCTCCC	TTCCTGGGCA	
CCGCCCCCC	CCCCCCCCC	00000000	TIGGCGGIGA	CGGIGIGGCA	TICCIGGGCA	60
	COGGCCCCGGI	GGTGCCGGCG	GGGCCGGTGG	GCTGTTCAGC	GTCGGTGGG	120
CCGGCGGCGC	CGGCGGAATC	GGATTGGTCG	GGAACAGCGG	TECCECCCC	MCCCCCCCC	
CCGCCCTGCT	CTCCCCCCAC	CCCCCCCCCC	222222	10000000	TCCGGCGGT	180
22222222	CIGGGGGGAC	GGCGGTGCCG	GCGGCGCGGG	TGGGGTCGGG	TCCACTACCG	240
GCGGTGCCGG	CGGGGCGGGC	GGCAACGCCA	GCCTGCTGGT	A A		
						282

- (2) INFORMATION FOR SEQ ID NO:262:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 415 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:

CGGCACGAGC	CGTGCTACTG	GTCAACTGAT	GCCCTGATTG	TGACCTTCCC	GGCGCCGGAT	
CAGTGCTTCT	CAGGACCGAC	GTAATATTCG	AAAACCAATC	CCCCCCCCCC	GGCGCCGGAI	60
AATGCCACAC	CGGCGGCGAT	CACCCACCC	AMARCCARIC	CGGCCGCCGA	GGCGAGGATG	120
CACCCCCACA) CCCCCCCAT	CAGCLACGGG	AGCCACAACG	CGATGCCGAC	CGCTGCCACC	180
GAGCCGGACA	ACGCGACCAT	GATCGGCCAC	CAGCTATGCG	GACTGAAGAA	TCCAAGTTCT	240
CCIGCGCCGT	CGCTGATTTC	AGCGCCTTCG	TAGTCCTCGG	GCCGGGAATC	TAACCCCCCC	300
GCCACAAACC	GGAAGAAGGT	GGCGACGATC	AACGCCATGC	CCCCCCCCC	CCCCCCCCCC	
ATGGTGCCAG	CCCACMCGAC	ACCACCCCCC	GCGAACATCG	CGCCGGIGAG	CGCCAACGCA	360
	acaici conc	WCCWCCOO I G	GCGAACATCG	AGGTCAACAC	GCCGT	415

- (2) INFORMATION FOR SEQ ID NO:263:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 373 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:

TATTGAGTTC CGGTTGTCAA	TGCCAGGG TGCCGTTGGA TACCAGTCTT TTCGAACGTT CGCGGCCTCG	ATCATCGGCT CGGTCCGTGG TGGGGGATAT GCGCAAGCGC	AGGCTCACGG CGTGGTGGGA GGCCTATTTG	TTTCGCCTGG AGTCTGACGC GTGTCGTCGG	GCCGCTCCAC	60 120 180 240 300 360 373
--------------------------	--	--	--	--	------------	--

(2) INFORMATION FOR SEQ ID NO:264:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 423 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:264:

CCAAACCGGA GCCAACCGGT GGCCGAGTAC CCACTCGTGC ATGATTGAGT TCTTGAGGGG AGAGGCCCGC	CCGCGGGCCG CGAATTCGAG TTAAACCGCT TTTTTTGCTG	AACCGCAGCG CTCCGTCGAC TAGCAATAAC	ATCCGTGGAT ACCAAAGCAA AAGCTTGCGG TAGCATAACC	CGCGCGTTCG CGCGATCGAT CCGCACTCGA CCTTGGGGCC	ACCCGGGTGA TCTAAACGGG	60 120 180 240 300 360 420
						423

- (2) INFORMATION FOR SEQ ID NO:265:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 404 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:265:

GGGTGCCGAC CAGAATGTCG CAGCACCCGC	GTCGTCACAT ACGTCGCGCA CTGCCGAGGC	CCGGCAGGCC GCGGCCGATC	AGCGATTTCG GGGTGCGGTC CCACCGGCTG	GCGAACCGGG GGATCGTGCT ACCAAGCGCT	CATCCGCCCA TATGCACCGC CGCCGTCCAG GTAACACAGC CGAGCACCTC	60 120 180 240
	CTOCCOMOGC	TITCGGGCGC		300003000		300
CAAACTCTCT	AGCGCCACCG	CGGTGCTGCG GACGGAACGT	GTGCGCGGCG	ACGGTCACCG	GAAAGTGCGA	360
			CACCCCGIII	GCGA		404

- (2) INFORMATION FOR SEQ ID NO: 266:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 421 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:266:

GTCCTGGTCG CAGGCTGT	IC TTCGAACCCG	CTGGCTAACT	TCGCACCCGG	GTATCCGCCC	60
ACCATCGAAC CCGCCCAA	CC GGCGGTGTCA	CCGCCTACTT	CGCAAGACCC	GGCCGGTGCA	120
GTGCGACCAC TGAGCGGC	CA CCCCCGGGCG	GCACTATTCG	ACAACGGCAC	CCGCCAATTG	180
GIGGCICIGC GCCCGGGC	C CGATTCGGCG	GCACCCGCCA	GCATCATGGT	CTTCGATGAC	240
ATGCACGTTG CACCGCGC	T CATTITICIG	CCGGGCCCGG	CAGCCGCGTT	GACCAGCGAC	300
GACCACGGCA CGGCCTTC	TT TGCCGCCCGC	GGCGGCTACT	TCGTGGCCGA	CCTGTCCTCC	360
GGTCACACCG CACGAGTG	A TGTCGCTGAC	GCAGCGCACA	CCGATTTCAC	CGCGATCGCC	420
С					421

(2) INFORMATION FOR SEQ ID NO:267:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 426 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:267:

ATGCATATCA CGCTCA	ACGC CATCCTGCGT	GCGATCTTCG	GGGCCGGCGG	CAGTGAACTA	60
GACGAGCTGC GCCGCC	ICAT TCCGCCGTGG	GTCACGCTGG	GCTCGCGCCT	GGCGGCGCTTA	120
CCGAAACCCA AACGCG	ACTA TGGCCGCCTT	AGCCCGTGGG	GCCGGCTGGC	CGAGTGGCGG	180
CGCCAGTACG ACACTG	TCAT CGACGAGCTC	ATCGAAGCCG	AGCGGGCCGA	CCCGAACTTC	240
GCCGATCGGA CCGACG	TITT GGCGTTGATG	CTGCGCAGCA	CTTACGACGA	CGGTTCCATC	300
ATGTCGCGCA AGGACAT	TTGG CGACGAACTG	CTCACGCTGC	TTGCCGCCGG	GCACGAAACC	360
ACGGCGGCGA CATGGG(ITGG GCGTTCGAAC	GGCTCAACCG	GCACCCCGAC	GTGCTCGCGG	420
CICIGG					426

- (2) INFORMATION FOR SEQ ID NO:268:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 522 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:268:

(GTCCTGGTCG	CAGGCTGTTC	TTCCNACCCC	GBGGGB			
		Chicacialic	TICGAACCCG	CTGGCTAACT	TCGCACCCGG	GTATCCGCCC	60
2	ACCATCGAAC	CCGCCCAACC	GGCGGTGTCA	CCGCCTACTT	CGCAAGACCC	GGCCCCTCCA	120
1	STECCACCAC	TCACCCCCA	55555555		COCIMONCCC	GGCCGGIGCA	120
,	aracancanc	IGAGCGGCCA	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	GCACTATTCG	ACAACGGCAC	CCGCCAATTG	180
(GTGGCTCTGC	GCCCGGGCGC	CGATTCGGCG	GCACCCGCCA	201 mas macm		
	-		carricacca	GCACCCGCCA	GCATCATGGT	CITCGATGAC	240
(STGCACGTTG	CACCGCGCGT	CATTTTTCTG	CCGGGCCCGG	CAGCCGCGTT	GACCACCCAC	200
(ZACCACGGCA	CCCCCTTCCT	TCCCCCCCC		aloccoco11	GACCAGCGAC	300
•	anc ancoder	COOCCIICCI	TGCCGCCCCCC	GGCGGCTACT	TCGTGGCCGA	CCTGTCCTCC	360
(GGTCACACCG	CACGAGTGAA	TGTCGCTGAC	GCAGCGCACA	CCCImmora	22222	
				GUAGCGCACA	CCGATTTCAC	CGCGATCGCC	420

CGCCGCTCCG ACGGCAAGCT GGTGCTCCCC ACCGGT TO TO	
CGCCGCTCCG ACGGCAAGCT GGTGCTGGGC AGCGCAGATG GCGCCGTCTA CACGCTTGCC AAGAACCCGC AGTTGACCGG CGTCGGCGCC GCCACCGTAG CC	48 0 522
(2) INFORMATION FOR SEQ ID NO:269:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 739 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:269:	
GCTGGGGCGC ACCGCCGTCC GGCGGCCCCA GCCCCTGGGC CCAGACCCCG CGCAAAACCA	
ACCCGTGGCC CTTAGTGGCC GGCGCCCCA GCCCCTGGGC CCAGACCCCG CGCAAAACCA TCGGCATCTG GATCGCCATC CGGCCCAAGG CCGTCGTGCT CGTCCTCGTG TTGGGCGCCA	60
TCGGCATCTG GATCGCCATC CGGCCCAAGC CGGTACAGCC GCCTCAGCCG GTTGCGGAGG AGCGCCTTAG CGCCCTACTG CTGAACTCCT GAAACTCCT GAAACTC	120
AGCGCCTTAG CGCCCTACTG CTGAACTCCT CAGAAGTCAA CGCCGTGATG GGCTCGTCGT CCATGCAGCC GGGCAAACCG ATCACATCGA TCCATGCAGCC GCCTGATG GGCTCGTCGT	180
CCATGCAGCC GGGCAAACCG ATCACATCGA TGGACTCTTC GCCGGTGATG GGCTCGTCGT CGGACTGCCA GGGCGCGCTG TATACCAGCC AGGATCGCGT CTCACATCGA GTGTCCCTGC	240
CGGACTGCCA GGGCGCGCTG TATACCAGCC AGGATCCGGT GTATGCCGGC ACCGGCTACA CCGCCATCAA CGGCTTGATT TCATCCGAGC GGGGGGAGATCAGA CCGGCTACA	300
CCGCCATCAA CGGCTTGATT TCATCCGAGC CGGGCGACAA CTACGAACAT TGGGTGAACC	360
AAGCCGTCGT CGCCTTTCCG ACCGCCGACA AAGCCCGCGC GTTCGTGCAG ACTTCGGCCG ACAAATGGAA GAACTGCGCA GGCAAGACGC TGAGGCGTGAAC ACTTCGGCCG	420 480
ACAAATGGAA GAACTGCGCA GGCAAGACGG TCACCGTCAC GAATAAGGCC AAGACCTACC GGTGGACGTT TGCCGACGTC AAAGGCAGCC CGCGGACGAC GAATAAGGCC AAGACCTACC	540
GGTGGACGTT TGCCGACGTC AAAGGCAGCC CGCCGACGAT CACGGTGATA GACACCCAAG AAGGCGCTGA GGGCTGGGAA TGCCAACGCC CGCCGACGAT CACGGTGATA GACACCCAAG	600
AAGGCGCTGA GGGCTGGGAA TGCCAACGCG CGATGAGCGT GGCCAACAAT GTGGTTGTCG ACGTCAACGC ATGCGGGTAC CAGATCACCA ATGCAGCGT GGCCAACAAT GTGGTTGTCG	660
ACGTCAACGC ATGCGGGTAC CAGATCACCA ATCAAGCAGG CCAGATCGCC GCCAAGATCT GTTGACAAAG TCAACAAGG	720
	739
(2) INFORMATION FOR SEQ ID NO:270:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 69 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:270:	
AGACGTCGTC GAGGCCGCCA TCGCCCGCGC CGAAGCCGTT AACCCGGCAC TGAACGCGTT	
GGCGTATGC CGAAGCCGTT AACCCGGCAC TGAACGCGTT	60 69
(2) INFORMATION FOR SEQ ID NO:271:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 523 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: CDNA	
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:271:	

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 224 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:272:

GTGTCGGTGT CGTCGGGGTA G CTGCAGGAAC CGGTGCCGGC G GCGCCTTGGC CGCGCCCTTG T AACGCTGAAG GGGCCCGGCG T	ICCACCGGGT	GGGCGACCAA	AGGCCGGATC		60 120 180 224
--	------------	------------	------------	--	-------------------------

(2) INFORMATION FOR SEQ ID NO:273:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 521 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:273:

TGAACTGACT	GCCCCGCTCG	ATCGGCGGCG	GCGGCGTGTC	1 m) acres		
TGAACTGCTC	TTCGCCATAG	CGGGCCTTCC	TCTCGGCCTT	ATAGCTGCGC	CGCCAGGCCA	60
CGTAGTGGCG	THE CHARLES	COGGCCTIGG	TOTOGGCCTT	GTCCAAACCC	TGCAGCGCGC	120
TGTCCNACCC	CACAMOC	CGCCAGCTAC	GCCGCACGGG	AATCCAGAGC	CGATCGGCGC	180
	CACATOCGCG	GIGGIGATCG	CCCCCCCCA	03 3 00 3 00 mm		240
	GICGIGIICC	GCGATCAGC		* * * * * * * * * * * * * * * * * * * *		
TGTCCGTCAG CGCCGTGGCG	GCCGACATCG	ACCCAGCCGG	TGAACACCTT	AACCGCCTCT	GCCIGGCCCT	300
CGCCGTGGCG	CAGCAACACC	AGGCTCCCAG	TGNACAGGII	GAGGGCATTC	CAGTCGCTCT	360
CGCCGTGGCG	CCTCATCCTC	AGGCTGCCAG	TGTTTGCCAT	ACCGGCAAGT	CTCTCACGCA	420
CTCCCGCACT	CCICMICGIG	GACCAAAATG	CCCGAATTCT	CCTCGGTCCG	CTGCGCAGCG	480
CGTTCATACC	GCCGAGGTGG	TCGGCACCGT	AACGGCCGGT	T		521

(2) INFORMATION FOR SEQ ID NO:274:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 426 base pairs
 - (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO:274: CTCCAGGCTC ATTCGCTCGA ACAAAGCCAC CCGGCCGTAC AGCGGACGCC CCCATTCGTT GTCGTGATAG TCGCGGTACA GCTGGGCATC GGGCCCTGGA CGAACCTCCG CCCAGGGGCA 60 GCGAACCAGC CCGTCGCCGC TCACGCGGGG TCAGAACGGT AGTGCACGAC AGTCTCGCCG 120 CGCGAAGGGT TTGACGCGTC AGACTCGGCC TCGGCGTCTT CCGACGAGGC GTGGATCGCC CCGAGCTGAG AGCGTAGCGC CTCGAGCTCA CGGCCGAGCC GTTCCAGCAC CCAGTCCACC 240 TCGCTGGTCT TGTTCCCGCG CAGCACCTGC GTGAACTTGA CCGCGTCGAC ATCGGCGCGG GTGACCCCGA ACGCCGGCAG CGTCGTCGCC GTCGTCGCCC GCGGCAGGGG CGGCAACTGC (2) INFORMATION FOR SEQ ID NO:275: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 219 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO:275: GCGGACACGG CGGACAAAGC GCAATCGGCC TCGGCGGCGG CGCCGGCGGC GACGGGGGCC AGGGCGGCGC CGGCCGCGA CTGTGGGGTA CTGGCGGCGC CGGCGGACAC GGCGGGGCAA 60 GGCGGTGGTA CCGGGGGCCC ACCGCTGCCC GGTCAGGCAG GCATGGGCGC CGCGGGTGGC 120 GCCGGTGGGC TGATCGGCAA CGGCGGGGCC GGCGGCGAC 180 219 (2) INFORMATION FOR SEQ ID NO:276: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 571 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA (xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:

AAGATCATCG GCGCCGCTCC TTAGCATCGC TGCGCTCTGC ATCGTCGCCG GCGCGGATCA CGGAGGTCCG GCCTTGTACC CCACTCCTCG AACGGTCAGC ACCACAGTCG GGTTCTCGGG ATCCTTTTCG ACCTTGGCCC GCAGACGCTG GACATGCACG TTCACCAGCC TGGTATCGGC 120 TGGGTGCCGG TAACCCCATA CCTGTTCGAG CAGCACATCA CGAGTAAACA CCTGGCGCGG 180 CTTGCGCGCC AATGCGACCA ACAGGTCGÃA TTCCAGCGGT GTCAACGAGA TCTGCTCACC 240 GTTGCGAGTG ACCTTGTGCG CCGGTACGTC GATTTCTACG TCGGCGATGG ACAGCATCTC 300 GGCGGGTTCG TCGTCGTTGC GGCGCAGCCG CGCCCGCACCC CGCGCAACCA GCTCCTTGGG CTTGAACGGC TTCATGATGT AGTCGTCGGC GCCCGACTCC AGACCCAGCA CCACATCCAC GGTGTCGGTC TTTGCGGTGA GCATCACGAT CGGAACACCCG GAATCGGCGC GCAACACCCG GCACACGTCG ATGCCGTTCA TACCGGGGCA A 571

- (2) INFORMATION FOR SEQ ID NO:277:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 93 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:277:

 Leu
 Phe
 Gly
 Ala
 Gly
 Gly
 Val
 Gly
 Val
 Gly
 Asp
 Gly
 Val
 Ala

 Phe
 Leu
 Gly
 Thr
 Ala
 Pro
 Gly
 Pro
 Gly
 Gly
 Ala
 G

- (2) INFORMATION FOR SEQ ID NO:278:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:278:

 Met
 Pro
 Pro
 Val
 Ser
 Ala
 Asn
 Ala
 Met
 Val
 Pro
 Ala
 His
 Ser
 Thr
 Pro

 1
 5
 10
 15
 15

 Pro
 Val
 Ala
 Asn
 Ile
 Glu
 Val
 Asn
 Thr
 Pro

 20
 25
 25

- (2) INFORMATION FOR SEQ ID NO:279:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:279:

- (2) INFORMATION FOR SEQ ID NO:280:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:280:

Trp Pro Ala Gly Arg Pro Met His Pro Ala Pro Gly Thr Ser Ala Asp

1 5 10 15

His Pro Pro Asn
20

- (2) INFORMATION FOR SEQ ID NO:281:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 140 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:281:

Val Leu Val Ala Gly Cys Ser Ser Asn Pro Leu Ala Asn Phe Ala Pro

1 10 15

Gly Tyr Pro Pro Thr Ile Glu Pro Ala Gln Pro Ala Val Ser Pro Pro 20 25 30

Thr Ser Gln Asp Pro Ala Gly Ala Val Arg Pro Leu Ser Gly His Pro 35 40 45

Arg Ala Ala Leu Phe Asp Asn Gly Thr Arg Gln Leu Val Ala Leu Arg
50 55 60

Pro Gly Ala Asp Ser Ala Ala Pro Ala Ser Ile Met Val Phe Asp Asp 65 70 75 80

Met His Val Ala Pro Arg Val Ile Phe Leu Pro Gly Pro Ala Ala Ala 85 90 95

Leu Thr Ser Asp Asp His Gly Thr Ala Phe Leu Ala Ala Arg Gly Gly
100 105 110

Tyr Phe Val Ala Asp Leu Ser Ser Gly His Thr Ala Arg Val Asn Val

Ala Asp Ala Ala His Thr Asp Phe Thr Ala Ile Ala 130 135 140

(2) INFORMATION FOR SEQ ID NO:282:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 142 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:282:

Met His Ile Thr Leu Asn Ala Ile Leu Arg Ala Ile Phe Gly Ala Gly Gly Ser Glu Leu Asp Glu Leu Arg Arg Leu Ile Pro Pro Trp Val Thr 25 Leu Gly Ser Arg Leu Ala Ala Leu Pro Lys Pro Lys Arg Asp Tyr Gly 40 Arg Leu Ser Pro Trp Gly Arg Leu Ala Glu Trp Arg Arg Gln Tyr Asp 55 Thr Val Ile Asp Glu Leu Ile Glu Ala Glu Arg Ala Asp Pro Asn Phe 70 Ala Asp Arg Thr Asp Val Leu Ala Leu Met Leu Arg Ser Thr Tyr Asp 90 Asp Gly Ser Ile Met Ser Arg Lys Asp Ile Gly Asp Glu Leu Leu Thr 105 Leu Leu Ala Ala Gly His Glu Thr Thr Ala Ala Thr Trp Ala Gly Arg 120 Ser Asn Gly Ser Thr Gly Thr Pro Thr Cys Ser Arg Leu Trp 135

- (2) INFORMATION FOR SEQ ID NO:283:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 163 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:283:

Tyr Phe Val Ala Asp Leu Ser Ser Gly His Thr Ala Arg Val Asn Val
115 - 120 - 120 - 125 - 125 - 125

Ala Asp Ala Ala His Thr Asp Phe Thr Ala Ile Ala Arg Arg Ser Asp
130 - 130 - 140 - 135 - 140

Gly Lys Leu Val Leu Gly Ser Ala Asp Gly Ala Val Tyr Thr Leu Ala
145 - 150 - 150 - 150 - 160

(2) INFORMATION FOR SEQ ID NO:284:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 240 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:284:

Trp Gly Ala Pro Pro Ser Gly Gly Pro Ser Pro Trp Ala Gln Thr Pro 1 Arg Lys Thr Asn Pro Trp Pro Leu Val Ala Gly Ala Ala Ala Val Val 25 Leu Val Leu Val Leu Gly Ala Ile Gly Ile Trp Ile Ala Ile Arg Pro 40 Lys Pro Val Gln Pro Pro Gln Pro Val Ala Glu Glu Arg Leu Ser Ala 55 Leu Leu Leu Asn Ser Ser Glu Val Asn Ala Val Met Gly Ser Ser Ser 70 Met Gln Pro Gly Lys Pro Ile Thr Ser Met Asp Ser Ser Pro Val Thr 90 Val Ser Leu Pro Asp Cys Gln Gly Ala Leu Tyr Thr Ser Gln Asp Pro 105 Val Tyr Ala Gly Thr Gly Tyr Thr Ala Ile Asn Gly Leu Ile Ser Ser 120 Glu ?ro Gly Asp Asn Tyr Glu His Trp Val Asn Gln Ala Val Val Ala 135 140 Phe Pro Thr Ala Asp Lys Ala Arg Ala Phe Val Gln Thr Ser Ala Asp 150 155 Lys Trp Lys Asn Cys Ala Gly Lys Thr Val Thr Val Thr Asn Lys Ala 165 170 Lys Thr Tyr Arg Trp Thr Phe Ala Asp Val Lys Gly Ser Pro Pro Thr 185 Ile Thr Val Ile Asp Thr Gln Glu Gly Ala Glu Gly Trp Glu Cys Gln 195 200 Arg Ala Met Ser Val Ala Asn Asn Val Val Val Asp Val Asn Ala Cys 215 Gly Tyr Gln Ile Thr Asn Gln Ala Gly Gln Ile Ala Ala Lys Ile Cys 230

- (2) INFORMATION FOR SEQ ID NO:285:
- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:285:

Asp Val Val Glu Ala Ala Ile Ala Arg Ala Glu Ala Val Asn Pro Ala 1 5 10 15 Leu Asn Ala Leu Ala Tyr

- (2) INFORMATION FOR SEQ ID NO:286:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 174 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:286:

Leu His Pro Ala Gly Ala Thr Asn Gly Ser Gly Gln Leu Ala Leu Pro 10 Val Glu Ala Pro Pro Arg Ser Val Pro Ser His Gly Glu Pro Leu Gly Ser Ala Ala Pro Glu Gly Leu Glu Gly Glu Phe Asp Asp Arg Ile Asp Glu Arg Phe Pro Val Phe Ser Ser Ala Ser Leu Ala Glu Ala Leu Pro Gly Pro Leu Thr Pro Met Thr Leu Asp Val Gln Leu Ser Gly Leu Arg Ala Ala Gly Arg Ala Met Gly Arg Val Leu Ala Leu Gly Gly Val Val 90 Ala Asp Glu Trp Glu Arg Arg Ala Ile Ala Val Phe Gly His Arg Pro 105 Tyr Ile Gly Val Ser Ala Asn Ile Val Ala Ala Ala Gln Leu Pro Gly 120 Trp Asp Ala Gln Ala Val Thr Arg Arg Ala Leu Gly Glu Gln Pro Gln 135 140 Val Thr Glu Leu Pro Phe Gly Arg Pro Gln Leu Ala Gly Gly Pro 150 155 Leu Gly Ser Val Ala Lys Val Val Val Thr Ala Arg Ser Leu

- (2) INFORMATION FOR SEQ ID NO:287:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 61 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:287:

- (2) INFORMATION FOR SEQ ID NO:288:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 133 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:288:

 Met
 Ala
 Asn
 Thr
 Gly
 Ser
 Leu
 Val
 Leu
 Leu
 Arg
 His
 Gly
 Glu
 Ser
 Asp

 Trp
 Asn
 Ala
 Leu
 Asn
 Leu
 Phe
 Thr
 Gly
 Trp
 Val
 Asp
 Val
 Gly
 Leu
 Thr

 Asp
 Lys
 Gly
 Gln
 Ala
 Glu
 Ala
 Val
 Arg
 Ser
 Gly
 Glu
 Leu
 Ile
 Ala
 Glu

 Asp
 Leu
 Leu
 Pro
 Asp
 Val
 Leu
 Tyr
 Thr
 Ser
 Leu
 Leu
 Arg
 A

- (2) INFORMATION FOR SEQ ID NO:289:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 63 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:289:

(2) INFORMATION FOR SEQ ID NO:290:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 72 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:290:

(2) INFORMATION FOR SEQ ID NO:291:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 74 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:291:

Pro Ile Pro Val Arg Ala Ala His His Glu 65 70

- (2) INFORMATION FOR SEQ ID NO:292:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 174 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:292:

CCGCACGTAA	CACCGTGAAT	TCAACCCAACC				
CGAACGGTTA	TTGACGCCC	TGAAGGGAGC	CGCTGGTCAT	GGGCCGATTC	TATCCGTGGG	60
TTTTCACGGCA	ACCAACCCCC	GGAGGCCACT	CCGCTGCCAC	CAAGTGGTGA	TATCCGTGGG	120
CICOOCA	ACGAACGGCG	GACACACCAC	TTGACATTCG	ACAGCACGGC	CGCG	174

- (2) INFORMATION FOR SEQ ID NO:293:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 404 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:293:

	TCGCAAACGG	GGTGACGTTC	COTTCCCCTCC				
	CCCTCCCCC	0010000110		CGCTAGAGAG	TTTGTCGCAC	TTTCCGGTGA	60
		GCWCCCCAGC	ACCGGTGAGC	سات کا تات کی ت	3 CMCC3 CCMC		120
	and to a contract	GAIGGCGCCC	GAAAGCCTCG	GCAGGCCCCC	CCECCCECEC		
	TGGTCAGCCG	GTGGGATCGG	CCCCTCCCCC	20000000001	GC1GGC1G1G	TTACAGCGCT GGCGAGCACG	180
	ATTCCGACCCC	100000000		ACGTCGACAT	TCTGCTGGAC	GGCGAGCACG	240
	····	Accidentific	CCGGATGTGA	CGACGTCGCC	N C C C C C C C C C C C C C C C C C C C	63.63.65.6	300
	COCCOMM	CGCTGCCA	CAGCCTGACT	CGGTGGCGGT	CACTTCCCCC	CATACCCGGT GATGGTCAGC	
•	TGACGTACCG	GGAGCTGGAT	GCATTGGCCG	1001000001	CAGIIGGG	GATGGTCAGC	360
			مصمة بالمهورون	ACCIGGCTGGC	CACT		404

- (2) INFORMATION FOR SEQ ID NO:294:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 134 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:294:

- (2) INFORMATION FOR SEQ ID NO:295:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 526 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:295:

GCTTCGACGG	CTACGAGTAC	CTGTTCTGGG	TCCCTTCTC			
CCAAGAAGAC	CACCAAGGCC	CTCCCCCC	TGGGTTGTGC	GGGCGCCTAC	GACGACAAGG	60
						120
						180
						240
AATATCGGCA	GCTGGGCGCC	2 A C 77 C 3 C C C	GCCCGCACIG	CITCAACACC	ATCGGCAAGG	300
TGGTGCGCGA	CAAGACCCTC	AACTACACCG	TGCTGCACCA	CACCCAGCTG	CTCAATCGGT	360
						420
					CTCATCCCTC	
CCGCGGGGGC	CACCTGAGCC	GAGATGCCGC	GCCATGCCCA	CCCCCC	CIGAICGGIG	480
			occurreccon.	CCGCAG		526

- (2) INFORMATION FOR SEQ ID NO:296:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 487 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:296:

CTCGCCGCCG ATCAGGCCGA TAGACGAACC						
CCCAGGAACA	TCGCCAACCC	ACCCATCCGA	TGGGAGGCAA	GAAAGACGGC	GCCGACAATG	180
CGCGGGTAGG	CGACGGCTCC	CAGGCGACTC	GGGGTAGGCG	TGACGTGCAC		240

480

540 600 610

CCGCGATAGG ACAGGGCGAG CAAGCCACCG GCAACGCCGG CCACATCGCT GGACACCTCG AGACCGTACT GCACCAACCT GAAGAGCTGA ACACTCGCCG AACGTGCAAC AGCTGCGAAC	420
MIIGG	480
	487
(2) INFORMATION FOR SEQ ID NO:297:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 528 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(5) 1010LOGI: linear	
(ii) MOLECULE TYPE: cDNA	
(xi) CECULENCE December 1	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:297:	
ACGAAGCGCG AGAATATGAG CCGGGGCAAC CCGGCATGTA CGAGCTTGAG TTCCCGGCGC	60
CTCAGCTGTC GTCGTCCGAC GGCCGTGGTC CGGCGTTGTA CGAGCTTGAG TTCCCGGCGC CCGACGCCGG CCATGCGATC CGGCTGCGCC CGGGGTTTGGT GCACGCTTTG GAAGGTTTCT	
CCGACGCCGG CCATGCGATC CGGCTGGCCG CCGCCCACCT CAAGGCGGCC CTGGACACAG	120
AGCTGGTCGC GTCCTTCGCG ATCGATGAAC TACTGGACTA CCGCTCGCGG CGGCCATTAA TGACTTTCAA GACCGATCAT TTCACCCACT CCGCTCGCGG CGGCCATTAA	180
TGACTTTCAA GACCGATCAT TTCACCCACT CCGATGATCC TGAGCTAAGC CTGTATGCGC TGCGCGACAG CATCGGCACC CCATTTCTCC TGAGCTAAGC CTGTATGCGC	240
TGCGCGACAG CATCGGCACC CCATTTCTGC TGCTGGCGGG TTTTGGAGCCG GACCTGAAGT GGGAGCGGTT CATCACCGCC GTCCGATTGC TGCGGCGGG TTTTGGAGCCG GACCTGAAGT	300
GGGAGCGGTT CATCACCGCC GTCCGATTGC TGCCGAGCG TTTTGGAGCCG GACCTGAAGT ATCGGCCTGG GCACCGTCCC GATGGCCGTT CCGCAGAGCC CCTGGGGTGTA CGGCAGAACC	360
ATCGGCCTGG GCACCGTCCC GATGGCCGTT CCGCACACAC GACCGATCAC GATGACCGCT CATTCCAACA ACCGGGAGCT ATCTCCGATT TTGALACTAC GACCGATCAC GATGACCGCT	420
CATTCCAACA ACCGGGAGCT ATCTCCGATT TTCAACCGTT CGATCTCC	480
	528
(2) INFORMATION FOR SEQ ID NO:298:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 610 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
COMM	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:298:	
CCAAGCCCGT CAAGGAGCCG GTGCCCCCCT TGGGGGGG	
CCAAGCCCGT CAAGGAGCCG GTGCCGGCCT TGCCTCCGGT GCCGCCGACG CCGGCGTTGC	60
TGGCCCCGCT GTTTAGGCCG TTTTCCCGG GGTTCCTAC GGTGCCGCCG CCCGGCAGCA	120
CGCCGTTGCC GCCGCTGGTG GCGTTCGGG CCCCGCCGTC ACCGGCTTTG CCGCCATCGC	180
CCCTTGACCC TTTGGCGGTG TCGATGCGG CCTGGTTGAC GTATTGTTCC ACCGGCCCGG	240
CCCTTGACCC TTTGGCGGTG TCGATCGCGG CGTCGATGGA TCCGCCGGACC ACGACGTGCG AAGCCTCGCC TGCCGCCGCA GCCGCCCAAC TCTGGTGATGA TCCGCCGACC ACGACGTGCG	300
AAGCCTCGCC TGCCGCCGCA GCCGCCCAAC TGTGTCGCGG CTCCTGCGAT TTGGCCCCGG	360
CCGACGAGAT GATGGGCACC ACCGGAGCCT GCGGCCGGTCT GGGGGAGGCC AGCGCGGGTT CGCGGTCACG CCATACGCGA CGGTGCGCG CCGCTTCGGA CATTTTGGA CATTTGGA CATTTTGGA CATTTTGGA CATTTTGGA CATTTGGA CATTTTGGA CATTTTGGA CATTTTGGA CATTTGGA CATTTTGGA CATTTGGA CATTTGA CATT	420

(2) INFORMATION FOR SEQ ID NO:299:

CGCGGTCACG CCATACGCGA CGGTGCGCCG CCGCTTCGGA GATTTGCAGG CTGCGTTGCA

CCAGATCGAG CAGCGGTGTG CCCAGGGACT GGGTTAGCCC GTTGGCGCCC CCGTTGTAGC

GGCGAGCGCA ATATCGGTGC CCACTCGACC CAACCGCGAC TCCATAAGCG ACACCATTCG

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 164 amino acids
 - (B) TYPE: amino acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:299:

Phe Asp Gly Tyr Glu Tyr Leu Phe Trp Val Gly Cys Ala Gly Ala Tyr Asp Asp Lys Ala Lys Lys Thr Thr Lys Ala Val Ala Glu Leu Phe Ala Val Ala Gly Val Lys Tyr Leu Val Leu Gly Ala Gly Glu Thr Cys Asn Gly Asp Ser Ala Arg Arg Ser Gly Asn Glu Phe Leu Phe Gln Gln Leu Ala Gln Gln Ala Val Glu Thr Leu Asp Gly Leu Phe Glu Gly Val Glu 70 Thr Val Asp Arg Lys Ile Val Val Thr Cys Pro His Cys Phe Asn Thr 85 Ile Gly Lys Glu Tyr Arg Gln Leu Gly Ala Asn Tyr Thr Val Leu His 105 His Thr Gln Leu Leu Asn Arg Leu Val Arg Asp Lys Arg Leu Val Pro 120 Val Thr Pro Val Ser Gln Asp Ile Thr Tyr His Asp Pro Cys Tyr Leu 135 140 Gly Arg His Asn Lys Val Tyr Glu Ala Pro Arg Glu Leu Ile Gly Ala Ala Gly Ala Thr

(2) INFORMATION FOR SEQ ID NO:300:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 161 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:300:

Leu Thr Gln Arg Asp Thr Gly Ala Ala Ile Gly Gln Gly Glu Gln Ala
Thr Gly Asn Ala Gly His Ile Ala Gly His Leu Glu Thr Val Leu His
130 - - - 150 - - - 140

Gln Pro Glu Glu Leu Asn Thr Arg Arg Thr Cys Asn Ser Cys Glu Gln
145 - - - 150 - - - - 150

(2) INFORMATION FOR SEQ ID NO:301:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 175 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:301:

Glu Ala Arg Glu Tyr Glu Pro Gly Gln Pro Gly Met Tyr Glu Leu Glu Phe Pro Ala Pro Gln Leu Ser Ser Ser Asp Gly Arg Gly Pro Val Leu 25 Val His Ala Leu Glu Gly Phe Ser Asp Ala Gly His Ala Ile Arg Leu 40 Ala Ala Ala His Leu Lys Ala Ala Leu Asp Thr Glu Leu Val Ala Ser 55 Phe Ala Ile Asp Glu Leu Leu Asp Tyr Arg Ser Arg Arg Pro Leu Met 70 Thr Phe Lys Thr Asp His Phe Thr His Ser Asp Asp Pro Glu Leu Ser 90 Leu Tyr Ala Leu Arg Asp Ser Ile Gly Thr Pro Phe Leu Leu Leu Ala 105 Gly Leu Glu Pro Asp Leu Lys Trp Glu Arg Phe Ile Thr Ala Val Arg 120 Leu Leu Ala Glu Arg Leu Gly Val Arg Gln Asn His Arg Pro Gly His 135 Arg Pro Asp Gly Arg Ser Ala His Thr Thr Asp His Asp Asp Arg Ser 150 155 Phe Gln Gln Pro Gly Ala Ile Ser Asp Phe Gln Pro Phe Asp Leu 165 170

(2) INFORMATION FOR SEQ ID NO:302:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 178 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:302:

Lys Pro Val Lys Glu Pro Val Pro Ala Leu Pro Pro Val Pro Pro Thr 1 Pro Ala Leu Pro Pro Leu Pro Pro Leu Pro Pro Val Pro Gly Phe Pro 20 25 Thr Val Pro Pro Pro Gly Ser Met Ala Pro Leu Phe Arg Pro Phe Ser 40 Pro Ala Pro Pro Ser Pro Ala Leu Pro Pro Ser Pro Pro Leu Pro Pro 55 Leu Val Gly Val Ala Ala Trp Leu Thr Tyr Cys Ser Thr Gly Pro Ala 70 Leu Asp Pro Leu Ala Val Ser Ile Ala Ala Ser Met Asp Pro Pro Thr 90 Thr Thr Cys Glu Ala Ser Pro Ala Ala Ala Ala Gln Leu Cys Arg 100 105 Gly Ser Cys Asp Leu Ala Pro Ala Asp Glu Met Met Gly Thr Thr Gly 115 120 Ala Cys Gly Arg Leu Gly Glu Ala Ser Ala Gly Ser Arg Ser Arg His 135 140 Thr Arg Arg Cys Ala Ala Ser Glu Ile Cys Arg Leu Arg Cys Thr 150 155 Arg Ser Ser Ser Gly Val Pro Arg Asp Trp Val Ser Pro Leu Ala Pro 165 170 Pro Leu

(2) INFORMATION FOR SEQ ID NO:303:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 921 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:303:

```					
AATTCGGCAC GARCAGO	LACC AACACCGGCT	TCTTCAACTC	CGGCGACGTC	AATACCGGTA	50
TCGGCAACAC CGGCAG	CTTC AACACCGGCA	GCTTCAATCC	GGGCGATTCC	AACACCGGGG	120
ATTTCAACCC ANGCAGO	ITAC CACACGGGGA	CTCGGAAACA	CCGGCGATTT	TACACCGGCS	180
CCTTCATCTC CGGCAGO	STAC AGCAACGGGT	CTTGTGGAGT	GGAAATTATC	AGGGCTCATT	240
GGNTGCACCC GGSCTT	RCGA ATCCCTCGKG	CCAATTCAAC	TCCTCNACAA	GCTTGCGGCC	300
GCACTCSAGC CCGGGTC	GAAT GATTGAGTTT	AACCGCTNAN	CAATAACTAG	CATAACCCCT	360
TKGGGCCTCT AAACGG	STCT TGAAGGGTTT	TTTGCTGAAA	GGANGAACTA	TATCCGGATA	420
ACTGGCGTAN TACGAA	AAGC CGCACCGATC	GCCTTCCCAA	CAGTTGCGCA	CCKGAATGGC	480
AATGGACCNC CCTKTT	ACCG GSCATTAACN	CGGGGGTGTN	GGKGTTACCC	CCACGTNACC	540
GCTACCTTGC CANNSS	CCTN RSGCCGTCTT	TCSTTTCTTC	CTTCCTTCTC	CCMCTTCGCC	600
GGTTCCCNTC AGCTCT	AAAT CGGGGNNCCC	TTTMGGGTTC	CAATTATTGC	TTACNGSCCC	660
CCACCCCAAA AAYTNA	TTNG GGTTAATGTC	CCTTMTTGGG	CNTCCCCCTA	WTNANNGTTT	720
TCCCCCTTNA CTTTGRS	STCC CTTCYTTATW	NTGAMNCTNT	TTCCACYGGA	AAAMNCTCCA	780
CCNTTYSSGS TTTCCT	ITGA WITATMRGGR	AATTSCAATY	CCGCYTTKGG	TTMAANTTAA	840
CYTATTICNA ATTITCO	CCGM TTTTMMNATR	TTNSNCKCGM	KNCTCCNRKA	SSGNTTTCCT	900
CCCCCYTTSS GKTYCC	CCRN G				921

## (2) INFORMATION FOR SEQ ID NO:304:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1082 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

### (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:304:

ACCGGGGTCA ACAGCACCAC GGTGGCGTCC ANGCAGAGCG CCGCGGTGAT GGCGGCCGAG 120  ACGGCRAACA CCTGCCGTAG CAGTCGGTGC GACTCCGCGC TCGCTCGANC CATGGCCGCG 180  CCGGCTGCCT CGAACANGCC TTCGTCGTCC ACAGCTTAGC CAGCANCCAA ACCGCACCCA 240  GAAACCCACA CGCCGCGCG CCCGGGANACC TGCGCCATCG KCTGCTGGGG CGANATCCCC 300  CGATCGCTNA CANGATGACC GCTGCCGGAA CGCCGCCGCT GCCTCCGGGC AGCCGCTGG 360  GCSGGGCAAC CGCGAACCCA NGAACACGGC AAGCAGTATC ANCGCAACAG CAATTGTCAA 420  GGGCTAAAACG CTTCACATCC AGGGATCTCG CGGCGCCACA CCGTCGGMTC TGCAGSGCGA 480  CCCCNTCCTN GGGCGGNCAC TCNTCAAAGA TGCNGATCNA CAGKCTAGGT CTTCGGCCGA 540  TATGSAAGGN CCCAACGGNT TTAAAGCCGC SAAAAAAASTC TCCCANTGGA TAAAATCAGC 600  CGGGGANCCC CAAAMTTAAG KTNGGGGNTY CGGGGCGGTA ACCGGCNNTK NGCCCCTTAA 720  AAAACCGGNC YTTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780  AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTC NTKGTTSCYT AWMCCCCCGG 840  AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTC NTKGTTSCYT AWMCCCCCGG 840  CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960  AMWTNYCCNC GWTTCAWTCS GGTCCGASMN AAACKCTTTY TTTTYCGSSC STCCMGGSNC 1020  SGGTKNANAN AAASATTMC YYCNNNANKK YYYCSSGCTT CYKMGRRNRR GMGAACCCGR 1080	330000000						
ACCGGGTCA ACAGCACCAC GGTGGCGTCC ANGCAGAGCG CCGCGGTGAT GGCGGCCGAG 120 ACGGCRAACA CCTGCCGTAG CAGTCGGTGC GACTCCGCGC TCGCTCGANC CATGGCCGCG 180 CCGGCTGCCT CGAACANGCC TCGTCGTCC ACAGCTTAGC CAGCANCCAA ACCGCACCCA 240 GAAACCCACA CGCCCGCGC CCCGGGANACC TGCGCCGCT GCCTCCGGGC CGANATCCCC 300 CCGATCGCTNA CANGATGACC GCTGCCGGAA CGCCGCGCT GCCTCCGGGC AGCCGCTGGGCGCGCGCGCGCT GCCTCCGGGC AGCCGCTGGGCGCGCGCGCGCGCGCGCGCT GCCTCCGGGC AGCCGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCT GCCTCCGGGC CAATTGTCAA 420 GGGCTAAACG CTTCACATCC AGGGATCTCG CGGCGCCACA CCGTCGGMTC TGCAGSGCGA 480 CCCCNTCCTN GGGCGGNCAC TCNTCAAAGA TGCNGATCNA CAGKCTAGGT CTTCGGCCGA 540 TATGSAAGGN CCCCAACGGNT TTAAAGCGGC SAAAAAASTC TCCCANTGGA TAAAATCAGC 660 CGGGGGANCCC CCCGTGSCMM NGTCYCGGKC ATTNTTCAAC MGGTTTNACG GCGGKTGCNG 660 GCCAACTKGC CAAAMTTAAG KTNGGGGNTY CGGGGCGGTA ACCGGCNNTK NGCCCCTTAA 720 AAAACCGGNC YTTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780 AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTC NTKGTTSCYT AWMCCCCCCGG AMCCYCCCGG GGKTGCNG CCCCCAWTGG CGGKTGKTC CANGNTYAAC 780 AAACCSGKYG GGKTGGCRTN WASSAMNCCC CMNGYYTCTT TAAAGGCCAN KNRAAWGKYT 900 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTTTYCGSSC STCCMGGSNC 1020 GGGGTKNANAN AAASATTTMC YYCNNNANKK YYYCSSGCTT CYKMGRRNRR GMGAACCCGR 1080		a	CGCACCGGGG	TCCGCAGCCG	GCGGGACCGT	CGCCAGCACC	60
ACGGCRAACA CCTGCCGTAG CAGTCGGTGC GACTCCGCGC TCGCTCGANC CATGGCCGCG 180 CCGGCTGCCT CGAACANGCC TTCGTCGTCC ACAGCTTAGC CAGCANCCAA ACCGCACCCA 240 GAAACCCACA CGCCCGCCGC CCCGGANACC TGCGCCGCT KCTGCTGGGG CGANATCCCC 300 CGATCGCTNA CANGATGACC GCTGCCGGAA CGCCGCGCT GCCTCCGGGC AGCCGCTGG 360 GCGGGCAACCCA NGAACACGGC AAGCAGTATC ANCGCAACAG CAATTGTCAA 420 GGGCTAAACG CTTCACATCC AGGGATCTCG CGGCGCCACA CCGTCGGMTC TGCAGSGCGA 480 CCCCCNTCCTN GGGCGGNCAC TCNTCAAAGA TGCNGATCNA CAGKCTAGGT CTTCGGCCGA 540 TATGSAAGGN CCCCAACGGNT TTAAAGCCGC SAAAAAASTC TCCCANTGGA TAAAATCAGC 600 CGGGGANCCC CCCGTGSCMM NGTCYCGGKC ATTNTTCAAC MGGTTTNACG GCGGKTGCNG 660 GCCAACTKGC CAAAMTTAAG KTNGGGGNTY CGGGGCGGTA ACCGGCNNTK NGCCCCTTAA 720 AAAACCGGNC YTTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780 AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTTC NTKGTTSCYT AWMCCCCCGG 840 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 900 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTTYCGSSC STCCMGGSNC 1020 GS	ACCGGGGTCA	ACAGCACCAC	GGTGGCGTCC				
CCGGCTGCCT CGAACANGCC TTCGTCGTCC ACAGCTTAGC CAGCANCCAA ACCGCACCCA 240 GAAACCCACA CGCCCGCCGC CCCGGANACC TGCGCCATCG KCTGCTGGGG CGANATCCCC 300 CGATCGCTNA CANGATGACC GCTGCCGGAA CGCCGCGCT GCCTCCGGGC AGCCGCTGG 360 GCGGGGCAAC CGCGAACCCA NGAACACGGC AAGCAGTATC ANCGCAACAG CAATTGTCAA 420 GGGCTAAACG CTTCACATCC AGGGATCTCG CGGCGCCACA CCGTCGGMTC TGCAGSGCGA 480 CCCCNTCCTN GGGCGGNCAC TCNTCAAAGA TGCNGATCNA CAGKCTAGGT CTTCGGCCGA 540 TATGSAAGGN CCCAACGGNT TTAAAGCCGC SAAAAAASTC TCCCANTGGA TAAAATCAGC 600 CGGGGANCCC CCCGTGSCMM NGTCYCGGKC ATTNTTCAAC MGGTTTNACG GCGGKTGCNG 660 GCCAACTKGC CAAAMTTAAG KTNGGGGNTY CGGGGCGGTA ACCGGCNNTK NGCCCCTTAA 720 AAAACCGGNC YTTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780 AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTTC NTKGTTSCYT AWMCCCCCGG 840 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTAAAGGCCAN KNRAAWGKYT 960 GS	ACGGCRAACA	CCTGCCGTAG	CAGTCGGTGC	GACTCCGCGC	_		
GAAACCCACA CGCCGGCGC CCCGGANACC TGCGCCATCG KCTGCTGGGG CGANATCCCC 300 CGATCGCTNA CANGATGACC GCTGCCGGAA CGCCGCGCT GCCTCCGGGC AGCCGCTGG 360 GCGGGGCAAC CGCGAACCCA NGAACACGGC AAGCAGTATC ANCGCAACAG CAATTGTCAA 420 GGGCTAAACG CTTCACATCC AGGGATCTCG CGGCGCCACA CCGTCGGMTC TGCAGSGCGA 480 CCCCNTCCTN GGGCGGNCAC TCNTCAAAGA TGCNGATCNA CAGKCTAGGT CTTCGGCCGA 540 TATGSAAGGN CCCAACGGNT TTAAAGCGGC SAAAAAASTC TCCCANTGGA TAAAATCAGC 600 CGGGGANCCC CCCGTGSCMM NGTCYCGGKC ATTNTTCAAC MGGTTTNACG GCGGKTGCNG 660 GCCAACTKGC CAAAMTTAAG KTNGGGGNTY CGGGGCGGTA ACCGGCNNTK NGCCCCTTAA 720 AAAACCGGNC YTTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780 AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTTC NTKGTTSCYT AWMCCCCCGG 840 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTAAAGGCCAN KNRAAWGKYT 960 GS	CCGGCTGCCT	CGAACANGCC	TTCGTCGTCC				
CGATCGCTNA CANGATGACC GCTGCCGGAA CGCCGCGCT GCCTCCGGGC AGCCGCTGG 360 GCSGGGCAAC CGCGAACCCA NGAACACGGC AAGCAGTATC ANCGCAACAG CAATTGTCAA 420 GGGCTAAACG CTTCACATCC AGGGATCTCG CGGCGCCACA CCGTCGGMTC TGCAGSGCGA 480 CCCCNTCCTN GGGCGGNCAC TCNTCAAAGA TGCNGATCNA CAGKCTAGGT CTTCGGCCGA 540 TATGSAAGGN CCCCAACGGNT TTAAAGCGGC SAAAAAASTC TCCCANTGGA TAAAATCAGC 600 GCCAACTKGC CAAAMTTAAG KTNGGGGNTY CGGGGCGGTA ACCGGCNNTK NGCCCCTTAA 720 AAAACCGGNC YTTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780 AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTTC NTKGTTSCYT AWMCCCCCGG 840 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRAAWGKYT 900 CCTTGGGAAW AAASATTTMC YYCNNNANKK YYYCSSGCTT CYKMGRRNRR GMGAACCCGR 1080 GS	GAAACCCACA	CGCCCGCCGC					
GCSGGGCAAC CGCGAACCCA NGAACACGGC AAGCAGTATC ANCGCAACAG CAATTGTCAA 420 GGGCTAAACG CTTCACATCC AGGGATCTCG CGGCGCCACA CCGTCGGMTC TGCAGSGCGA 480 CCCCNTCCTN GGGCGGNCAC TCNTCAAAGA TGCNGATCNA CAGKCTAGGT CTTCGGCCGA 540 TATGSAAGGN CCCCAACGGNT TTAAAGCGGC SAAAAAASTC TCCCANTGGA TAAAATCAGC 600 CGGGGANCCC CCCGTGSCMM NGTCYCGGKC ATTNTTCAAC MGGTTTNACG GCGGKTGCNG 660 GCCAACTKGC CAAAMTTAAG KTNGGGGNTY CGGGGCGGTA ACCGGCNNTK NGCCCCTTAA 720 AAAACCGGNC YTTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780 AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTTC NTKGTTSCYT AWMCCCCCGG 840 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRAAWGKYT 900 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTTTYCGSSC STCCMGGSNC 1020 GS	CGATCGCTNA				_		300
GGGCTAAACG CTTCACATCC AGGGATCTCG CGGCGCACA CCGTCGGMTC TGCAGSGCGA 480 CCCCNTCCTN GGGCGGNCAC TCNTCAAAGA TGCNGATCNA CAGKCTAGGT CTTCGGCCGA 540 TATGSAAGGN CCCAACGGNT TTAAAGCGGC SAAAAAASTC TCCCANTGGA TAAAATCAGC 600 CGGGGANCCC CCCGTGSCMM NGTCYCGGKC ATTNTTCAAC MGGTTTNACG GCGGKTGCNG 660 GCCAACTKGC CAAAMTTAAG KTNGGGGNTY CGGGGCGGTA ACCGGCNNTK NGCCCCTTAA 720 AAAACCGGNC YTTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780 AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTTC NTKGTTSCYT AWMCCCCCGG 840 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTTTYCGSSC STCCMGGSNC 1020 GS	CCCCCCTTC	- <del>-</del>	*******			AGCCGCGTGG	360
CCCCNTCCTN GGGCGGNCAC TCNTCAAAGA TGCNGATCNA CAGKCTAGGT CTTCGGCCGA 540 TATGSAAGGN CCCAACGGNT TTAAAGCGGC SAAAAAASTC TCCCANTGGA TAAAATCAGC 600 CGGGGANCCC CCCGTGSCMM NGTCYCGGKC ATTNTTCAAC MGGTTTNACG GCGGKTGCNG 660 GCCAACTKGC CAAAMTTAAG KTNGGGGNTY CGGGGCGGTA ACCGGCNNTK NGCCCCTTAA 720 AAAACCGGNC YTTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780 AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTTC NTKGTTSCYT AWMCCCCCGG 840 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRAAWGKYT 900 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTTTYCGSSC STCCMGGSNC 1020 GS CCCCCARTG CCGCCAACTCC CTYMMGSSCN CTTKCWRTYN NRNGGGAACCCGR 1080						CAATTGTCAA	420
CCCCNTCCTN GGGCGGNCAC TCNTCAAAGA TGCNGATCNA CAGKCTAGGT CTTCGGCCGA 540 TATGSAAGGN CCCAACGGNT TTAAAGCGGC SAAAAAASTC TCCCANTGGA TAAAATCAGC 600 CGGGGANCCC CCCGTGSCMM NGTCYCGGKC ATTNTTCAAC MGGTTTNACG GCGGKTGCNG 660 GCCAACTKGC CAAAMTTAAG KTNGGGGNTY CGGGGCGGTA ACCGGCNNTK NGCCCCTTAA 720 AAAACCGGNC YTTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780 AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTTC NTKGTTSCYT AWMCCCCCGG 840 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTTTYCGSSC STCCMGGSNC 1020 GS		CTTCACATCC	AGGGATCTCG	CGGCGCCACA	CCGTCGGMTC	TGCAGSGCGA	480
TATGSAAGGN CCCAACGGNT TTAAAGCGGC SAAAAAASTC TCCCANTGGA TAAAATCAGC 600 CGGGGANCCC CCCGTGSCMM NGTCYCGGKC ATTNTTCAAC MGGTTTNACG GCGGKTGCNG 660 GCCAACTKGC CAAAMTTAAG KTNGGGGNTY CGGGGCGGTA ACCGGCNNTK NGCCCCTTAA 720 AAAACCGGNC YTTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780 AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTTC NTKGTTSCYT AWMCCCCCGG 840 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTTTYCGSSC STCCMGGSNC 1020 GS	CCCCNTCCTN	GGGCGGNCAC					
CGGGGANCCC CCCGTGSCMM NGTCYCGGKC ATTNTTCAAC MGGTTTNACG GCGKTGCNG 660 GCCAACTKGC CAAAMTTAAG KTNGGGGNTY CGGGGCGGTA ACCGGCNNTK NGCCCCTTAA 720 AAAACCGGNC YTTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780 AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTC NTKGTTSCYT AWMCCCCCGG 840 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTTTYCGSSC STCCMGGSNC 1020 GS	TATGSAAGGN	CCCAACGGNT					
GCCAACTKGC CAAAMTTAAG KTNGGGGNTY CGGGGCGGTA ACCGGCNNTK NGCCCCTTAA 720 AAAACCGGNC YTTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780 AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTTC NTKGTTSCYT AWMCCCCCGG 840 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTTTYCGSSC STCCMGGSNC 1020 GS CYKMGRRNRR GMGAACCGR 1080	CGGGGANCCC		<del>-</del>				600
AAAACCGGCC YTTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780 AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTTC NTKGTTSCYT AWMCCCCCGG 840 AAACCSGKYG GGKTGGCRTN WASSAMNCCC CMNGYYTCTT TAAAAGGCCAN KNRAAWGKYT 900 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTTTYCGSSC STCCMGGSNC 1020 SGGTKNANAN AAASATTTMC YYCNNNANKK YYYCSSGCTT CYKMGRRNRR GMGAACCCGR 1080				ATTNTTCAAC	MGGTTTNACG	GCGGKTGCNG	660
AAAACCGGCC YTTCTKGAT TAMMACCGGN CCCCCAWTGG CGGKTGKTCC CANGNTYAAC 780 AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTTC NTKGTTSCYT AWMCCCCCGG 840 AAACCSGKYG GGKTGGCRTN WASSAMNCCC CMNGYYTCTT TAAAGGCCAN KNRAAWGKYT 900 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTTTYCGSSC STCCMGGSNC 1020 GS CYKMGRRNRR GMGAACCGR 1080		:	KTNGGGGNTY	CGGGGCGGTA	ACCGGCNNTK	NGCCCCTTAA	720
AMCCYCCCSS MNGGGKTGGS SAACCCTTCC CGNGGGGTTC NTKGTTSCYT AWMCCCCCGG 840 AAACCSGKYG GGKTGGCRTN WASSAMNCCC CMNGYYTCTT TAAAGGCCAN KNRAAWGKYT 900 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTTTYCGSSC STCCMGGSNC 1020 SGGTKNANAN AAASATTTMC YYCNNNANKK YYYCSSGCTT CYKMGRRNRR GMGAACCCGR 1080	AAAACCGGNC	YTTTCTKGAT	TAMMACCGGN	CCCCCAWTGG	CGGKTGKTCC	CANGNTYAAC	. – •
AAACCSGKYG GGKTGGCRTN WASSAMNCCC CMNGYYTCTT TAAAGGCCAN KNRAAWGKYT 900 CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTTTYCGSSC STCCMGGSNC 1020 SGGTKNANAN AAASATTTMC YYCNNNANKK YYYCSSGCTT CYKMGRRNRR GMGAACCCGR 1080	AMCCYCCCSS	MNGGGKTGGS	SAACCCTTCC	CGNGGGGTTC	MAKCHARCAA		
CCTTGGGAAW CCTNCAATYC GAAAAYYCTC CTYMMGSSCN CTTKCWRTYN NRNGGGAACS 960 AMWTNYCCNC GWTTCAWTCG GGTCCGASMN AAACKCTTTY TTTTYCGSSC STCCMGGSNC 1020 SGGTKNANAN AAASATTTMC YYCNNNANKK YYYCSSGCTT CYKMGRRNRR GMGAACCCGR 1080	AAACCSGKYG	GGKTGGCRTN	WASSAMMOOC		<del>-</del>		
AMWINYCCNC GWITCAWICE GGICCGASMN AAACKCITTY TITTYCGSSC SICCMGGSNC 1020 SGGIKNANAN AAASATITMC YYCNNNANKK YYYCSSGCIT CYKMGRRNRR GMGAACCCGR 1080 GS	CCAMCCC 22M				TAAAGGCCAN	KNRAAWGKYT	900
SGGTKNANAN AAASATTIMC YYCNNNANKK YYYCSSGCTT CYKMGRRNRR GMGAACCCGR 1080					CTTKCWRTYN	NRNGGGAACS	960
SGGTKNANAN AAASATTIMC YYCNNNANKK YYYCSSGCTT CYKMGRRNRR GMGAACCCGR 1080	AMMINYCONC	GWTTCAWTCG	GGTCCGASMN	AAACKCTTTY	TTTTYCGSSC	STCCMGGSNC	1020
GS 1990	SGGTKNANAN				CAKMCSBNDD		
1082	GS				C T TO TO TO TAKE	GHAMCCCCK	
							1082

### (2) INFORMATION FOR SEQ ID NO:305:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 990 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:305:

AATTGGCACG	AGTGATCGCG	CTGAAGCCGG	TAGCGCGGGT	GGCTCGGGTG	GTTTGCGAAC	60
RAAATCCGCT	CGANGTGGTC	TCGGTAGGCG	GTGTCCANAA	CCCTCCCCCC	COCCOCCO	
א יוייייר א יוייריר	CCCCCCCC	***************************************	GIGICCANAA	600100666	GTGCCGGCGG	120
MICIONICOG	CGCGGCCGTA	GTGCACGTCG	GCGGGCGTGT	GCAGTCCGAT	GCCGGAATGC	180
TTGTGTTCGT	GGTTGTACCA	GCCGAACAAC	CCCMCCC3 cm			
C3.000003.500		CCCGMAGMAC	COG T COCHCI.	GCACCCGGGC	CGCCTCGATC	240
GACTCGAACC	GTTTCGGGAA	ATCGGGCCGG	TACTTGAAGG	TCTYCAACTC	GGCCTCAGAC	300
AACGGGTTGT	CTTGCTGGTG	TCCCCCCCC	1.555.55	·	GGCCTCAGAC	300
	CTTGCTGGTG	raceacce.TG	AGTGCGACTT	GGTGACACCG	AAGTCGGCCA	360
NCANCAATGC	CACCGGTTTG	GAACTCATCC	ACAACCCCCC	TCCCCCTCMA	CCTC3 CTTCT	430
NCGCCGCTAA	THE PROPERTY OF THE PROPERTY O	00022		I C C G C G I C MA	GGICACTIGI	420
IICOGCGC IAM	TTTNYTGGGC	GGCAAGGGTT	TGCCGAYCAN	KCCGCTCGGC	CAAAACTTCG	480
ANTCNCSCCA	AGGCCNCCAT	CCNCCCAAAC	3 MCmm3 CCCC	3 3 7 3 3 3 3 3 7 3 7 7 7 7 7 7 7 7 7 7		•••
Casan Co Co Cosant	TTD TO LOCATE		MING LACGGG	ANAAAANATY	CAAAGAYCAC	540
CYTCCGGKTN	LIATANCTYC	CCYTTTGSTY	GGGCCCCCCN	CYYTGKKNAT	ACCCCMNCCa	600

ANTCGGAAMC AKGCCCCCNY CKYYSSMYCC YKGCCCCCC	ACTWMMNACC CGGNTSTACC TCSGCNWNNC CCCCTCCCCM AMMNNNGGGG	CNNGGGSCCY AAAAASCCCK CSGCGGKKKT CTCCGNKTCC	AAMCGTYYNR CCNWTCCCTC KKGTTNCCCT CCAMCCYANC CCCCMKRRGM	AGGITTISCT CRASATTGSC WMRCWMWYTS	WTAATTTAAA NAAAGAAASA NCCSAAWKSA GGCCNASCCN GKKCCCWKNT MCCTCWGNRA	660 720 780 840 900 960
		THE CHICKETTAIN				990

## (2) INFORMATION FOR SEQ ID NO:306:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 223 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:306:

AATTCGGGTG	GCAACGCGGG	CCTGTTCGGC	770000000		CGGTGGGGCT	
GGTGGTGGCG	CCCCCCCCC	22222	MACGGCGGCG	CCGGTGGTGC	CGGTGGGGCT	60
001001000	CCGGCGGCGC	GGGCGGTAAC	GCGGGGTGGT	TTGGTCATGG	GGGCGCTGGC	120
GGCGTGGGTG	GTGTANGTGC	GGCCGGGGCC	AACCCCCCC	22222222	GGATGGGGCG	120
CCTCCTCTTC	CCCCCTCCC	22.00000CC	MCGGIGCIA	CGCCCGGTCA	GGATGGGGCG	180
4616616116	CCGGGICGGA	CRACRCTCGT	GCCGCTCGTG	CCG		223

## (2) INFORMATION FOR SEQ ID NO:307:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 418 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:307:

AATTCGGCAC	GANGCGGCAA GTGCCGGCGG	CGGTGGCAGC	GGCGGCACGT	CNGTTGCCAC	CGGGGGGCC	60
GGGAACGGCG GGCGGCAGTG	GCGGAATGGG	CGATCCCCC	GGGGCCGGGC	TGATCGGCAA	CGGCSGCAAC	120
GGCGGCAGTG (	GTTTGGACPG	CGCCAACCCC	GGCGGCACCG	GCGTCNGCGG	CATCRGTGGG	180
CTGTTGTTGG (CAGCACAGGC (CGGCAACGGC)	GTTGGCCGCA	GTCAACCCC	CCGGCCAGCA	CCAACCCGCT	GCACACCGCG	240
CGGCAACGCG	CCAACGGCGC	CCCCCCCAAC	CCATCCAGGC	CGTGACCGGG	CGCCCCTGAT	300
TTCGGCGGCG	GAAGGAACGG	CGGGTCCCCC	GGGGCCCCCC	GCRGGCACGG	CGGGTGGTTG	360
		COGGICCGGC	GICANCRGCG	GGGCGGCGG	AAATGCCG	418

## (2) INFORMATION FOR SEQ ID NO:308:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1049 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO:308:

AATTCGGCAC GA	GGGGCACG ATCGCATA	Ch ccccmana			
GCTCGGCACA CG	CGAGCGCA CANTAGG	CA GCGCTCGCGG	CAGACCCGCC	CGATACAGCA	60
TACCGGCCAC CA	CGAGCGCA CAATACGG	CG TCTGGCTGTC	CGGCTTGARC	ACCACCGCGT	120
GCGAGATCAC CC		CG ACACCGTAAG	CGTCATGGGG	TAGTTCCACG	180
		TT GATAGCACAC	CGTGGTCTTG	CCTATCCCGG	240
GCAGCAGCGG CT	GTGCCTTA CGGGGCTT	CA GCAGGTCCAC	ACAGACTCGT	GCSTTATAAT	
TNCGCSTTCC GC	GAICAGAT CGACAATT	TC CTCTTCCCC	CCCCAMCCCC	0000000	300
CTCGGCTTGC AG	GAAGTCCA TGAAGAAC	TC GCGCTTCTCC	ATTINACACO		360
GCSGATGACT GC	AGCTCGCT CGATNACC	CC VCCCCCCC	ATMAACAGGT	CGCGATAGCG	420
CTTCCGCGAA TG		GG ACCTTCGCCA	GTCGGTCTGC	GCCGCGCGAN	480
		GG NCGTGCCAAC	GGAATCNTAT	CACGGGTTGC	540
AACSANNCAA ACC	CTCAATST NCYGGTCG CTCGGCAA GGTTAGGM	AA ATTCGGCAAC	TTCTTATCCC	GGCAGGTRCC	600
CAN AMERICAN ACC	CICGGCAA GGTTAGGM	IT TCCCCCNCTT	YCAAAAATNC	GGKTTTTGGN	660
	CNATGKTG MCAAGGMT	CT CKAANAAKCS	GGGTCYTCTN	NTCNGKGGAK	720
CCAAAMGGKT TTO	egggmagc gknmncca	AN CCTWACCCTG	KTKAANGGNW		
GGGAKKGNGA ATY	YCYCCSNA NCCCRGGG		TYCCGGMCTC		780
WGMGSTTTCC CAP	AAAAACSC CCCAAATT			CTCKGGAWTC	840
MMCSSAARNS ANM		TK AAAAAGNAYW		CTTTTKARCA	900
	NCTNTTT TSCNMTWC			TYTAWTTSSC	960
	YNTKGYN KNTATMAGO	M WNYTNCRMCC	MMMSNCKSNG	KKGGNRCCNN	1020
	TITING IN KNTATMAG	-			1049
					-

# (2) INFORMATION FOR SEQ ID NO:309:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1036 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:309:

AATTCGGCAC			AATGGTGAAG	CCTCGGTGCC	TGCCGTTACG	50
CCAAGAKTCA		CCCCCCGGTG	GGAATGCTGA			
AGGGCTGGGG	TGGAATAACT	GAANGTTACT		ACCCGGTATT		120
GGGCCGATCA	ANGTTGTGGG				GATATGTATT	180
TTGTYRTGGA	CRAKACWGGC		GATGGTTTGG		ATTTGGGGAA	240
CANGGTGATG				TTSAANTTTT	GTGCCGSCCA	300
CGAGATYGCC			SATCGAAATA		CNACGCCSAA	360
		TGGGCGGGAC	AACCMASGGT	CCSANGTAAK	GGTTTCCTTN	420
ATNTTGATCG		- TO TO TO TO	TGSGCTCSAY	MTSATSGCCC	NACNCCWCCG	480
YTTATTTCMS	GCTNAYGGGA	ATBAMRGGAA	CAAYNTCCCT	CCCMGGAAAA	ACCHACMECC	
CCTGGTNSYC	CNCCCRCCNC	AKAACCCRTT	KCTGTPSTMC	CCSMAAATNA		540
NACTCCNCSG	AANTNSCCCC	CCCSCKNNTT	ATSTYCCCGK	CCSMAAATNA		600
TCCCCGGTTA	ACCCCCWTNT				CCCTTNAAMC	660
	SAMCWNCCNC		YTAAKMNCRG		CCCCCYTRMK	720
KTTNTMCTKC			CCCKCYKGSM	TNCCCAATNT	WCMWCKCCNS	780
	- CILITINGIC			WTATAAAACC	WCWYAWYNNK	840
KCNCWMAWTA		NCCCCNCNCK	NTTKTAMWCC	CKMCCCKCSW	TWCYCKCSCC	900
CCMTCTMNAC	YCCCCCKKTY	NKWMCCCTTC		MCNMBMKTCT	YCSGKTWCWC	
NCYNTIMICN	CYNANMCKCK	TTTI CTT	CRNTCTCCCC			960
CNCNCTCCSC				CCMCCCCCC	KKCTCTSKCC	1020
	_					1036

- (2) INFORMATION FOR SEQ ID NO:310:
- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1036 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:310:

AATTCGGCAC	GAGATCATGA	ATAGCGGGCT	CCTCACCACC	CN N CTCCTCC	0001mcmccc	
GAGCAAGTCT	CGTCTGCTCG					60
	· - <del>-</del>			GATGCGGACA	CCTGCGATGT	120
		AGGTAAGGCC		CTTTGCTAGC	AGGGTGTCTT	180
GGCTCTTCGC	ACGTGAGGTA	ACCAATAACT	CCGACGCAGA	CCAACTCCGG	CCCTCGATCC	240
GGGTACCAGG	CTCCGCCGGA	GCCAGCCGTT	GTGCCCCCTG	GGCCGAAGGT	CAGCTGCTGT	300
GCGATCGAAG	TAAGAAACCG	CGCCATGCCC	GTCGCCAAGT	ACGACTGACC	GAGCAAACGA	360
ACGATCGTCG	TCCTTTCCGT	GGGGGTAATC		ACCGCACGAG		420
TTGGGATTCG	GCCACTGACC	GACCAACCGC		CCCCAGCGGA		480
TTCCGCGGGG	CCGCNAACGG	AATCANCGSG				
ACATANCAAC	GGNNTCTGCG			CGAASCANCC		540
		CCCACATTTC	GGGSTTMTGC	CCCTCNGCAA	CSSNAAYNCC	600
CCCAATTCYG	AACNAAAAA	TTGGYCCATY	ARNGTYCTCM	CCAAAAACCN	AWTCCCCKTA	660
TCCCCCGGGG	GGGRCCCCYY	NMNAAAACGG		CCSGGGCSCC		720
CCCCTTGTCG	GCCCNCCSGG	TTTGGTCMCM		GGGNTGCSCC		780
AAAAAYCKNG	NCAAATYAAA	CCCKYCMAAA				
ANTTAANCCN	KAAAAAAAWW					840
				GGKYTTAGGG		900
ARAAAATMTC	CANATMINSSK	TTNNAAAAAA	ASCCSWAKCC	CCCNNNKKNN	CCAAWKAARR	960
SRCCTTCGGG	TNWNSGGGGG	KKKKKTNCMS	KMNMMTTWGR	CCCNCCGCCN	NNTWKCCTTN	1020
TCCNYGGNGC	RNCAGN					
						1036

### (2) INFORMATION FOR SEQ ID NO:311:

### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1060 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:311:

AATTCGGCAC	GAGTCGATTC	GATCGAACAC	GCCCGCACCT	GGCCAGGCCA	CATGGGCGCG	60
GCCATGGCCA	ACGCCTACTC	GGCCAACCCG	AATCCATTCG	GCGTCTCACC	GCAACCCCCG	120
AAACCGGCGA	CCGCGGCATG	GATCAACCCG	CCCACCCCAG	ATCCGAAATA	GCGTCCACAT	180
AATGAGACAC	TGGCGCAAAG	AGCTTGACAG	GCGCCGCACC	ACGCAAGCTG	TTAGACGTGT	240
CGGTCTTGCA	AGAAGCGGGT	TGGCCACCCA	AGATCACGCC	GCCCAAGGGC	ATCGAGTCAA	300
CGTTGCGGTG	GTATCGCGCT	AACGTCGGCG	CCGCCAAGAA	ATGACGGTGC	GCATTACCAT	360
GGCCCTGCTG	ATCACCTTTG	GCCACCTGCG	CACCANAACT	ATGANCAGCC	TTATECCEAG	420
TCTCGTGGAC	ATCGGCAGCC	GCTTCAAAAA	CTCCTTGTCG	ACAATSGTAT	TECTGANCCG	480
CCGAATTCTT	NTRCTTGCAA	SAACACTNCA	TGTTNCSGGT	NAACAACCYT	CCTTMCAAA	540
ACANCCAATA	TTGAANTCCC	ANTCGGGCAM	GAACCNGTTM	CGGAAGETGE	TCCCAACCAA	500
TGKTGCCCAA	AAATCCCGGG	NGGTRAAAWW	CCCNSNATCG	MENTATION	CTNCAACGAA	660
AAAAGGTCCA	AGKYCAAAGG	NGCCCCCCC	SCHANATEC	TCDACCCAVA	CINGAACAAM	
WWWTNCAAAT	MTTNGGGTCC	KNNTCCCCWT	AAANCCCCC	CCCMICMCC		720
NWNMGGGMGN					GMGTYTCCCC	780
	CCGGKGGGTM	AAAAAAAMMM CAAAAAAANAN	ACCCCCCC			840
	CC0011000111	CUMUMANIAN	ACCCCCCAMS	NGGGGGGAAA	ATTTGNAAWT	900

AAGGKKKTKC SCMACCCCAA AAANMMNNCN AWNCCCGMGK SARGGGGRNY TTMKAGGGMG 960 GNYCCCCCCW YCGGGGGGNA NAAYAAAAGK NGSNGRGAAT NTTNTTTTGK RSSSRNKTTT 1020 TYNTCCTYCN CCNMGNRWWG SRAMNTGKTS NSSGGGSGGC 1060

- (2) INFORMATION FOR SEQ ID NO:312:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1040 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:312:

AATTCGGCAC	GAGCTTCACC	AAAGAGCTGA	СУТСССССТ	GATGCGACAT	0003.0003.00	
GCAATACGGG		CCGAANGGAN				60
CCAAGGTGAA	<del>-</del>	GCGAAAGATG		GCTCAACTGG		120
TGTTNGTATG						180
		CCGCGCTGAC	NGATAANGAA	TTCGCTGGTC	GCCGGGCACN	240
ATGGATGGTC		TCCGCSGTTA	AATTGCSTGT	GCATCATCTG	GCAGGCTATG	300
TTCCCGCTAC	RCTGCAGCCC	ATCATGGATG	TGCGGCTAAC	GAANAAGTTA	TGACATGGCG	360
CAAGCGAMTC	GGGCATSCNC	GCGGCAMTTT	CGCAACCTGC			420
CGAATGCGGC	GCTYAAAAGC	NGGCTTGCGT	TGATTMMAAC			
CCGNGNMNTG	CGTTCTCTCC	AACTCCGKKG	SYTGCCNCCG		CNATYCTTTG	480
GTTGGACTTA	MRTNTTCAAA			TGAAACCCMA	CTNCCCCCCC	540
SAANTCGGGC	TTYGGGNRCC			SAACCTNCCR	TCAAANTAMM	600
		CCCCNGAAYW	TTCKNCNGGG	GMNNTYCTCN	GGTTYNGGCG	660
SAAACNTTTG	CCRTNCYMNN	TTTACAMGGC	NCMINMITGM	GGGSCSNNAS	GWCCCGGGKK	720
TNTTTNCAAW	TCNCNSKTTT	TTKGGGGGGG	GGCYGRTRMC	NCGGGCCCCC	GGCCCKKMAA	780
AAAAAMCMSA	RRCCNCYGGG	KKCCCCCCCM	NNATNGGGCG	YKCRAAACAA	ACCCCAAMBA	840
TNGNGMGGGC	SMACCSGNGN		TSNSCTMANM	MKGMANNNCT	SGMSCCMNSN	
NCTGMGGGKT	TTKGNNGARN		RCGGNCGCNN			900
NGASNGWMGN				GAAAGGGSMS	GSCKSCNNGN	960
	CGYMTNKCGC	MCMGIGNMRN	NNGNNNGNNN	GGGRKNNACN	NMKMCAWSMC	1020
	COTHINKCGC					1040

- (2) INFORMATION FOR SEQ ID NO:313:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 348 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:313:

AATTCGGCAC	GAGACAANGG	CGTGAAATGG	GATCCGGCCG	AGCTGGGGGC	CGTCGTCAGC	<b>C</b> 0
GACCTGTTGG	CCAAGTCGCG	CCCCCCCCC		AGC 1GGGGCC	COLCOLCAGO	60
3.000001100	CCANGICGCG	GCCGC CGGT-F	CCGGTCTATG	GGGCCTAGTT	ATCTGCGCCG	120
AGCGTGAACT	CAGGGCGAGA	TTTCGGCCGT	TTTCTCGCCC	TGGCTTCACG	TTCGGCGAAG	180
TKGGGAACGG	TCAGGGTTCG	CAAACCACGA	TOCOCTOCO	20000000000	CAGGACTGGT	
AMERCAGATA	בט עשטטאששט	TOOMOLOGA	1CGGGATCGT	GCGGTCGGTC	CAGGACTGGT	240
ANTECIGATA	CIINGGIACA	TUGTGACCAA	CTGTGGNCAA	TATTCGGCGC	GCTCCTCGTC	300
NGTCGCGTCC	CGCGCGGTAA	GGTCCANCAC	TTCCTTTTTC	TOGTGOOG		340

(2) INFORMATION FOR SEQ ID NO:314:

	CS:	ISTI	RACTERI	Έ	SEQUENCE	i)	1
--	-----	------	---------	---	----------	----	---

- (A) LENGTH: 332 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

#### (ii) MOLECULE TYPE: Genomic DNA

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:314:

AATTCGGCAC	GAGAGACCGG	GTCGTTGACC	AACGGACGCT	TGGGCGCGGG	CCCCTTGCGT	60
GGCATCAGCC	CITCTCCTTC	TTAGCGCCGT	AACGGCTGCG	TGCCTGTTTG	CGGTTCTTGA	120
CACCCTGCGT	ATCCAGCGAA	CCGCGGATGA	TCTTGTAGCG	CACACCAGGC	AGGTCCTTCA	180
CCCGGCCGCC	GCGCACCAGC	ACCATCGAGT	GCTCCTGCAG	GTTGTGGCCC	TCGCCGGGAA	240
TGTACGCCGT	GACCTCGAAC	TGACTCGTCA	CTTCACGCGG	GCAACCTTCC	GAAGCGCCGA	300
GTTCGGCTTC	TTCGGAGTGG	TGGCTCGTGC	CG			332

#### (2) INFORMATION FOR SEQ ID NO:315:

#### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 962 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

#### (ii) MOLECULE TYPE: Genomic DNA

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:315:

AATTCGGCAC	RAGTCGGTCT	AGACGGATTC	AATGCTCCCG	CGAGCACCTC	GCCACTGCAC	60
ACCCTGCAGC	AAAATGTGCT	CAATGTGGTG	AACGAGCCCT	TCCAGACGCT	CACCGGCCGC	120
	GCAACGGCGC	CAACGGGACT			GGGGCCGGCG	180
GGTGGCTGTT	CGGCAACGGC	GGCAACGGCG	GGTCCGGGGC	GAACGGAACC	AACGGCGGGG	240
ACGTGGGGAC	GCGCCCGGCG	GGATTTCTTC	GCACCGGSGC	ACCGGCGGG	CCGGCGGCGT	300
CGCACAACGG	CACCGGCGGG	GACGCNGCGC	CCGTNGGGCG	GCTTCTKGAT	GGGCTCCGGC	360
GGTNACGCGG	CACGGCGGCG	CCCGGCTCAC	CGCCNGTTGG	GACGCGGGGA	CGCGTNACCC	420
CGATCTTCTT	CCGCNCCCCG	GAAACCGCGG	GGCCGGCCCC	ACATTAKACC	CGGCGGNACC	480
GCGGMCCCGG	CGGAACGGNG	GGYNTTTTCC	AACGGCGGGG	CCGCGGAACC	GNMGGSTGTT	540
CCTTNGGSGA	AGGNCCAAKT	CCCGKCTANC	YYAATCCCCG	ANGGKTGAMC	CTSATGSNCA	600
MYTTMAGGAA	CYTNCCCANT	KTTSGRACCW	CRCCNGGAAA	ASRAWNKNGT	KGGCAAACNA	660
NNTNCYTTKN	NATTKGGNNA	AAAANCCCTY	CCWCSGRACT	NCCCCCCNGM	GRGMCNNTNN	720
NTTTYGNCNN	CCCGGSNAAM	RNTTKATTTC	NGGGGGNTCN	GGGTKMNNNA	AACCCCAAAM	780
MNRNNKCSCA	ANGGGKSNGC	NKNNMMNSGT	TTTYCKNMRA	MRNWTYKNKN	NTCNGARSRN	840
NAAMCNNSNK	NGKKKNNKAA	ARNNTTWKTN	KNSCNNNCNN	GRRNGVRGGC	CKMKGSNMNG	900
MCWHNAWRNG	NNGSNCNCKC	NNKMNAAAAA	AASGGVNCKS	NSMKNKKKKG	NRGGGGGGG	960
GG						962

#### (2) INFORMATION FOR SEQ ID NO:316:

#### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 323 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:316:

AATTCGGCAC	RAGAAGACGC	CCGAANGTTT	GCGCTGGCTC	TACAACTTCA	TCAARGCGCA	60
GGGGGAACGC	AACTTCGGCA	AGATCTACGT	TCGCTTCCCC	GAAGCGGTCT	CGATGCGCCA	120
GTACCTCGGC	GCACCGCACG	GCGAGCTGAC	CCAGGATCCG	GCCGCGAAAC	GGCTTGCGTT	180
GCAGAAGATG	TCGTTCGAGG	TGGCCTGGAG	GATTTTGCAN	GCGACGCCNG	TGACCGCGAC	240
GGGTTTKGTG	TCCGCACTGC	TGCTCACCAC	CCGCGGCACC	GCGTTGACCT	CGACCAGCTG	300
CACCACTCGT	GCCGCTCGTG	CCG				323

- (2) INFORMATION FOR SEQ ID NO:317:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1034 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:317:

AATTCGCAGT	GTGTGTGGCG	GCGTCCAGAA	GAAGATGATC	GCGAACATCG	CCAGCGCCGG	60
CCAGGCTATG	GTGCCGGTGA	TGGCCGACCA	GCCGATCATC	ACCGGCATAC	AGCCGGCCGC	120
CCCACCCCAC	ACCACGTTCT	GTGACGTGCG	TCGCTTGAGC	CAAAGCGTGT	AGACRAACAC	180
ATAAAACGCG	ACGGTGACCA	GGGCCAGCAC	CCCCGCCAGC	AGGTTCGTGG	CGCACCATAG	240
CCAGAAGAAC	GAGATCACCG	TCNACGTCAC	CCGAGTGCCA	ACGCGTTTCG	GGTCGGCACC	300
GCTTCCCGCG	CCAAGGGCCG	GCGCGCGGTT	CGCTTCATCA	CCTTGTCGAT	ATCGGCGTCG	360
GCNACCAGTT	GAGCGTGTTG	GCGCCGGCGG	CSGCCATCAT	CCCGCCGACN	ANCGTGTTGA	420
GCATGANCAG	CGGATGAATG	GCGCCGCGGC	TCGTGCCGCT	CGTGCCGAAT	TCAACTCCGT	480
CNACAACTTG	CGGNCGCACT	CGAACCCGGG	TGAATGAWTG	AATTTAAACC	GSTSAACANT	540
AACTACATAA	CCCTTGGGGG	CTCTTAACCG	GTYYTGAANG	GGTTTTTTGC	TTAAAGGAAG	500
AACYATTTCC	GGATANCTGG	CSTTNWTARC	GAAAAGGCCC	CRCCCATNGC	CCTCCACAGT	660
TTSCCCCTGA	ATGGSAATGG	MNCNCCYKNR		- · · · · · · · · · · ·	GGNTTTTGKT	720
MCCCNNCTKA	CNTTMMMTGC	ARNNCNGGCC			NTCCCCCNST	780
TNCNGKTCCC	CNNAMNYTNW	ACGGGGGGCC	YTNGGGKCRM	TWTKKTTTGG	GCCCCMCCCC	840
MAAANASAAN	GGGGKRNGTY	CSTTTGGCNC		NYCCCCCCAM	YTNRRKMCSY	900
CNNTNKGGNN	CTGTNCKNCG	GAARAMAMCC	KCCCCGNSTS	STTNGTYWAG	GNRWKGNSRG	960
CCSCCCCGGY	MNNNAAYAWN	WMNATNCNNS	STNANMAKKN	NNNNNNSCN	WNGNGNNTCN	1020
SCNSNGGKBC	CCCC		~		WINGTIGHTAL CIA	
	COCC					1034

- (2) INFORMATION FOR SEQ ID NO:318:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 331 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:318:

AATTCGGCAC GAGCCCACAT RAGGCACAGT CGCTGGAGCG TCCCGGGTGG TGTTGACCCA GACGGGTCGC CGCGGTGATC TGCTGCGCAG TGAGCGCTCG CGGGCCGCGC TGAGTGCGCC	CAATGTGTTG CGACATCGCC GAGAAGCTCA CCGATCGCCG	CTGACCGTGC CAGCGCGACA AAGGTCATCC CGCTGGTCAC	TGTCCCGGTT ACCTGCGGGT GTTGTTCGCC	GGGGACCGGT CGGCCGCCAC	60 120 180 240 300
CGGGCCGCGC TGAGTGCGCC	TCCCGCGAGC	A			331

- (2) INFORMATION FOR SEQ ID NO:319:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1026 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:319:

AATTCGGCAC	GAGATCGTCA	CCCTGGCGAC	CAGTGCACCC	AGGCCACGCC	ACCAGTTACG	60
GCTGATGGGC	CAGAAGATGG	ACCAGGTGCT				120
CACCGGGATC	GCGGTCCTCA	GCTACGGCGA			CCGCTGACTA	180
TGACGCCGCG	TCCGAAATGC	AGCAGCTGGT	CAACGGTATC		TGGCGCGTCT	240
GGTGGCGCTC	ANCGACAATT	CCGTGCTGCT	GTTTACAAGG	ATCGGCSTAA		
CGCGCACTCC	CCANCGCCGC	GCGGCSGGG		TGCCGACCGC		300
CACTCACCCC				· · ·	CCGAGCGCGT	360
CACTGACGCC	ATCTCCGTCG	GCGTTAACCC		TGGGTCGTGC	GCAAGTTGGG	420
CCCGGTCACC	ATCNATCCGC	GCCGCCATGA	CGCNGTGCTG	TTCCACACCA	CNTSNGACNC	480
CCCCCAGGAA	CTGGTCCGGC	AMTNCAGGAA	NTYCGTGTGG	GCACCNGCTT	CTTCCGKTRT	540
GGCYTAAACT	TCCNATSTTN	CSGCSGGCCT			NTCTTNCCAA	600
ATCGGSMMAA	ATCCCCANMC	333000000				800
		AAACCCCCCG	GGTCTTGSGG	GCSGGGNGGC	GGCCNAWNCC	660
AAACCCCCC	NTTAAANTCT	TTGKTNCCNN	CNCSGGCNCC	NCNAANSCAN	CCCTTTKGGC	720
NCTTCCCCCC	CCCAWTTTAA	CCGAKCGSCN	AAYCCCAAGY	TMMGKCCYCY	KNAAAAAAAA	780
AATTTGSCSG	CCCCAANTAA		CCYTTGGGGG			
				CGRANCNYNT	TTTMCCSNSS	840
	NGGANCESGG	KAAYTMMTKG	NAAYCGCCSN	AAMBNTTTTC	TAANNCCCCN	900
YNCCCSGAAA	ATTNNAMAAM	CMNNKTGSNG	GGGGKTTSNC	SGKKGRAGGM	AAAAAANRSN	960
SKTTNMCNNN	SANMNCNSNN	SGGNSNNNNN	NNNCNCGYKC	CSNAANMCCC	000000000	
CCMMCC	· ··		THE THE GIVE	COMMANMICCC	CGCGGGGGGG	1020
CCATICC						1026

- (2) INFORMATION FOR SEQ ID NO:320:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 324 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:320:

AATTCGGCAC	GAGAAGACGC	CCGARNGTST	GCGCTGGCTC	TACAACTTCA	TCAARGCGCA	60
NGGGGAACGC	AACTTCGGCA	AGATCTACGT	TCGCTTCCCC	GAAGCGGTCT	CGATGCGCCA	
GTACCTCGGC	GCACCGCACG	GCGAGCTGAC	CCAGGATCCC	CCCCCCAAAC	GGCTTGCGTT	120
GCAGAAGATG	TCGTTCGAGG	TECCETCEN	CLAUGAICEG	GCCGCGAAAC	TNACCGCGAC	180
GGGTTTKGTG	TCCGC3CTCC	TOGECTOGAN	GATTTTGCAN	GCGACGCCNG	TNACCGCGAC	240
9691117919	i cederat i Ge	TGC TCACCAC	CCGCSGCACC	GCGTTGACGC	TCGACCAGCT	300

#### GCACCACTCG TGCCGCTCGT GCCG

324

- (2) INFORMATION FOR SEQ ID NO:321:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1010 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:321:

3300000000						
AATTCGGCAC	GANGCGTGCC	GCTNAACACC	AGCCCGCGGC	TGCCAGATAT	CCCGGACTCG	60
GTAGTGCCGC	CGGTGGCGTC	GTTGCTCTCC	TGACGGGGCG	CGGCGACCAT	AAGGTCGCTM	120
ATGCCCAGGT	AGCGGCCCAG	GTGCATGGAG	TCGATGATGA	TGCGACTCTC	CAGCTCGCCG	180
ACCGGGAGCT	TGGCATCGGG	CCTGATCAGC	CAGGACGCGT	AGGACAAGTC	GATCGAATGC	240
ATAGTGGCCT	CCAGAGTGGC	CGTGCAMTTC		CACGGCAAAT		300
CTACTCCGCG	TANTGTTCCC	GCATCGCCTG		GGGAACCGCA		360
	GANCTCAGGT	TTGCCGCTTT				
				GTCNACANCC	GGTACTCGGC	420
ATANATCTGG		GCGCCGACGG		AANAACGGGC		480
CCGCCCCGGT	CACCCNAACA	ACANCTTGSC	ATCGGATTTT	GTCCCCANCG	CTCAANCCGT	540
CCCGAACGCC		NACTITICIT		CCGCTTCCGK		600
WTAAATGGGA	AACCCTTNCC	CCACCTTGAA			GSTAACCCCG	660
	GANTCGGTCN					
	GWILLCGGICM	KCCGGGSTTT	YSTNTTCCCC	ACCTINGNAN	GGGCCGGCCA	720
AGSTTTTCTT	SYTGAAGGGG	GAAACCCAAC	TTTNTYTYYN	AACCSCMNAA	MYMTTTYCSG	780
MNAASCCNKT	CCCCTTTAAC	CAMGGSGGTN	AACCGKTMNG	NGGKTAAAAA	GGGSKNNKTG	840
NCCCCYMANG	GGGGGRAAAA	TSTKTCNNCG	GGGCCKAAAW	ACCMMMMYGN	GTGKKKNKSS	900
GCSAAATTTT	NMMRAACTKN	000000000				
		GGGGCCSSGA		MSCCCCCSNN	GSTGKCCCNN	960
NTTTCCNNAA	WMKKGKNWNM	SNMNSCSNGG	GKYNSGGSNN	NNAAGMGGGG		1010

- (2) INFORMATION FOR SEQ ID NO:322:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1010 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:322:

AATTCGGCAC	GANGCGTGCC	GCTNAACACC	AGCCCGCGGC	TGCCAGATAT	CCCGGACTCG	60
GTAGTGCCGC	CGGTGGCGTC	GTTGCTCTCC	TGACGGGGCG	CGGCGACCAT	AAGGTCGCTM	120
ATGCCCAGGT	AGCGGCCCAG	GTGCATGGAG	TCGATGATGA	TGCGACTCTC	CAGCTCGCCG	180
ACCGGGAGCT	TGGCATCGGG	CCTGATCAGC	CAGGACGCGT	AGGACAAGTC	GATCGAATGC	240
ATAGTGGCCT	CCAGAGTGGC	CGTGCAMTTC	CNGCGTGCTC	CACGGCAAAT	GCCTTGATTT	300
CTACTCCGCG	TANTGTTCCC	GCATCGCCTG	CGGGATGAAT	GGGAACCGCA	SGATGGCGAC	360
		TTGCCGCTTT		GTCNACANCC	GGTACTCGGC	420
		GCGCCGACGG		AANAACGGGC	ACNACAATCG	480
CCGCCCCGGT	CACCCNAACA	ACANCTTGSC	ATCGGATTTT	GTCCCCANCG	CTCAANCCGT	540
CCCGAACGCC	TCNTCCGGCG	NACTTTTCTT	NNAWTAACTG	CCGCTTCCGK	CCCTGGNGCA	600
WTAAATGGGA	AACCCTTNCC	CCACCTTGAA	GGGGTTGTTG	NATTTTTACT	GSTAACCCCG	660

AATTNTTCCG GANTO	GGTCN KCCGGGSTTT	YSTNTTCCCC	ACCTINGNAN	GGGCCGGCCA	720
AGSTTTTCTT SYTGA	LAGGGG GAAACCCAAC	TTTNTYTYYN	AACCSCMNAA	MYMTTTYCSG	780
MNAASCCNKT CCCCT	TTAAC CAMGGSGGTN	AACCGKTMNG	NGGKTAAAAA	GGGSKNNKTG	840
NCCCCYMANG GGGGG	RAAAA TSTKTCNNCG	GGGCCKAAAW	ACCIMIMIMYGN	GTGKKKNKSS	900
GCSAAATTTT NIMMRA	ACTION GGGGCCSSGA	NNTTTNAAAG	MSCCCCCSNN	GSTGKCCCNN	960
NTTTCCNNAA WMKKG	KNWNM SNMNSCSNGG	GKYNSGGSNN	NNAAGMGGGG		1010

- (2) INFORMATION FOR SEQ ID NO:323:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1092 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:323:

NGNGGGGWNS	NTCAYCAYCA	YCACSGGGYW	CWATTGCGGC	CGCAWCTTGT	MAASAGATCT	60
CGAAYTCGGC	AMGAGGGAMT	CKCTMGCNCC	GCTGTGCAAN	CCAATRAGGC	CTRATAATTY	120
CCACTCCACA	AAAAACCGTT	GTGTGTAYYT	SCCGRAAATR	AAGGCGCCGG	TNTCAACWYC	180
GCCGGTKTTY	CCRATYCCCG	TKTTGTAMCT	GCCKGGGTSR	AAAYCCCCGG	TGTTGGAYCC	240
CCGGATTGAA	ACTGCCGGKT	TGAAACTGCC	GKTTTSGCSA	TCCGGKWATT	GAMSTCRCGG	300
ATTAAAAAAC	CGGKKTTGGN	GCTGSNCGTG	CCAAATNCGR	AYCCRATAYC	CCATGGCCTG	360
KYCTYCTCCK	YCGGTACCCA	AAYCTGGGTA	TCCTATACTG	GYCCCTAAAK	GCAAWYCKGG	420
GCTGYCMMTK	TTGCKGGSGT	CCNAATTTAS	CACCASCGGT	TCCTTCCATA	CCNAAACNCG	480
CKTGGGCWCC	AGMCCGRAAA	AAAKAATAAT		CATNYCCAAA		540
					CTMMAACCCC	600
CAAANCCCCA	TAACNTNCGR	GAASAAACCC				660
TTGCTKSTTT	CGGGMWCCGT		YCCCAAASTA		CCNAGAKAAA	720
ACCNCGGGCN	CCMCCCSNAA	NWTATYTCTT	KGGCAANCCC			780
ATRMTCCCTT	CCCCVSCAAT	TGGYCGGRAT	NCGSNCCYTY		CAKWWNNGNG	
GRRNNACCMA	ACCCCAAGTY	CCMNAAAATN	GKCCCCGCTC	CNAACACGNK		840
		AACCCCCCNA	RKANTNCCCA			900
			NMCCCCGKKK		GGCCCCCCC	960
				KKTTTTCTTT	TKCCMRSCCC	1020
	CS	GGAAGRANC.V	TYCCSANANM	TCCCNYWRSW	CCGSWGMGNA	1080
						1092

- (2) INFORMATION FOR SEQ ID NO:324:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1251 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:324:

GGGGGGGNNN NATAC	CATCWT CYGTGYACC	G GGGMTCTAKT	GGCGGGCCGC	AATCTNGTCA	60
ASAGATCTCT NAMT	TCGGGC ACAAAAACT	W GACAAASYMT	CGNGCNMTCC	GTGTCCTNKA	
TCGCAAAACG NGTRA	ACASAC ASACACRTA	T GTGTGCCCAC	CASCAAYTCK	TTGGGACCTC	180
GCTRACCGGY TGCCC	CRNACS CCACGYTGC	S CWTCTATCCC	RACGCCGGCC	ACGGGYGGGG	240

ATATTCCAGG	CACCACGCCC	AGTTTGGTGG	ACAATGCCCT	GGCAKTTTCC	TCRAANTTCG	300
TGAAACCGAA	TTCNSMTTGA	ACCNCCAARG	CCCCSNCCNR	AACARTTGGG	WTCCGCGGTT	360
CTCCCCACCG	KTTTCCGGGG		AANCGCACCC			420
GGGCGGACAA	NTCGGGTTGC	AATTTTGCRA	AYCGGGGCCG			480
GAAACTGTTY	YCRAAMACCG		TTTCCGGGCR			540
GCTTRTACTT	CCCCGACCGT		ATCGTCNTNN		TGGGGCAGGG	
CKAAAYACCG	CMTTKGGTTT		GGCCCAANTC			600
GGNTCGAATT	SCCCCCGGT		NTGGCCNNYT			660
KGGCNSCCCC	AGTAANACCC					720
GGNTTCCGGK	· <del>-</del>					780
NASSKAYCCS					NCSTRTGKCA	840
	GGAAKAATWT			NSSGCCCCCC	AWAYGNKSTG	900
GGGGGKTTTT				NTSTGGAMCS	SGGACYCCCR	960
YTYYCYCCIM	TCCCCCNCSA			NTGNSGNGGG	KWNTTTATTT	1020
	TKACMSGGGG		GGGGGAGAAA	ANAAAAAAA	RAKGGYKNTT	1080
		NAGAGKTCCT	CKCKCCNCSG	SNITTCITTI	MGNSGSYGGG	1140
GNNGNNNAAA				NCNGGGGYGS	NGSCGNSTYN	1200
GNNKGRKWTA	TNIMGNCGTN	SCCTCCNCCC	GCKNKNTGTC	TMTCNMYGSG	C	1251

### (2) INFORMATION FOR SEQ ID NO:325:

#### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1099 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

#### (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:325:

AAYTCGGCAC	MGAGTATCAC	CAAKCTGYGT	GGCCCAGCAA	AGTGGAGCTA	TTACTACCTG	60
TATGTGATCC	TCRACATCTY		KTGGTCGGGT		CTCGCKTGAK	120
TCRAAGGTCT	TGGCCRAACG	GCTGATCGCG	CAAACCCTTG	CGCCCAGCAC	ATCAKCGCCG	180
AACAGCTGAC	CTGCMCGCCG	ACCGGGGGYC	GNCAATAACT	CCAAACCGGT	GGCMCTGCTG	240
CTGGCCNACY	CCGTGTCCCA	ANTCGAACTC	ASCCSGCNMA	CCAKMAACKA		300
CTGAAGCCCA	GTTCAAAAAC	CTCAAGTWCC	GGCCCRACTT	CCCGAAACGG	TNCGAGTCKA	360
TCRSAGGSGG	CCGGGTGCMC	TGCAACCGGT	TCTTCGGNTG	GTRCAMCCCN	AAAMCAAGCA	420
TTCCGGGMTC	CGMMTGCCCA	CGCCGCCAAS				480
GGGAACSGSN	CCMCCKTCNK	GGAMACGCCC	TWCCAAAACC			540
NAACNCCCGA	RCNCCCKSKT		NMSGCGAATA		CCGAATCCAA	600
TTCCCMKYGG	CITITYYYCC			CCTASSNMKC	KNCCAMNANT	660
CCNWATCTGG	NGGTCCCNAN	KYYGGCGTTC	NMAATSAMNA		TSCYACCMMN	720
AACCGKNNKG	KCCCCMKCTK	MANAAAKATT			NAAMACCSCN	780
CNCYNCWYTC	TMYCSSKWGC	GCSMYNANCA	SNGGGGAGGW	GGSGRMKMCT	CTMTCTCNCT	
MGCGCCKNTN		ACASMNKTCC	GCGCNGCGCN	MAAMANRAKA	CTAKCCGYGN	840
CCSNSTMTYN	CTSNNMKMNN	TCCWMWNATC	NTYYGKKCNN	KCTMKATNWC	CSCTSKCNCK	900
MRAMTCKTYG	SNMTCCTCCA	TCNCTCKKSC		KSCNCCNCWN		960
			WACNCACACK		CNKCNMKCWN	1020
	MIMICWCCS		MACHICACACA	NGWC1111CC	WKNNMKCNKM	1080
						1099

#### (2) INFORMATION FOR SEQ ID NO:326:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 296 base pairs
  - (B) TYPE: nucleic acid

258

- (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:326:

GNGNTATACA	TCWCTGTGYA	CCSAGGATCW	ANTGCGGCCG	MAAKCTWSTM	CASAGATCTC	60
AAAYTCTGCA	MGAGCGGCAC	AKAKYSTCGT	CCMRACCCGG	CAYACWCCWG	CNCGCCCCWT	120
CTTRGACCGG	GGCKATASMC	ACCGTTGGCC	CCGGCNCGCA	CCTACACCAC	CCACGCCGCC	180
AGCGCCCCCW	TRAMCAAACC	ACCCCGCKTT	TACCGCCCGC	GCCGCCGGG	CCACCACCAG	240
CCCCACCGGC	ACCACCGGCG	CCGCCGTTGC	CAAAACAGGC	CCGCKTTTGC	CACCRA	296

- (2) INFORMATION FOR SEQ ID NO:327:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1073 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:327:

NGNGSGNKMY	ATCATCWTTC	TGCACCSNGG	MTCWATTGCG	GCCGCAATCT	TSTMNASAGA	60
TCTCGAAYTC	GGCAMGARCA	TCTGCGCGGN	GAATGTCCAA	AWGTCWKTAA	CGGCMATCGG	120
TTTGCCGYCA	ACCACKCTRT	SCAKATGCGG	GCCAMWTYCA	AACCRATTAT	TTGGGYCGAG	180
AAAATTTMCG	CKTGTRASCA	ACCTGCAGCG	GGTCAASCAA	CAGCCTCTRA	ACCGTAAATY	240
CKTAGGTNKT	YCCGGCAACA	ASCYCRATAA	TSCGGCCCGC	AMCCACAAAA	CCTGANTNGT	300
TNTTCNCRAA	NCCGGTYCCC	GRAGGGGTSA	ACTGCSGTAR	GCTTNTCWYC	NCCTTRACAT	360
TAAACCCCCC	CGGNTCWTCG	CCGCGCCCAA	ATYCYTGCCC	WTKGCNACCA	YCCCANCCTG	420
CSGTATGGTS	RAANCASTSG	GCRAACGGTM	MCCSTACCKC	TGGCTGATYC	KTCGGNTCCS	480
SNAATTCGGG	GATTTACGGS	CAMGGTTAAY	CCAGGYCCCC	TNTGCYTCKY	CNACAACCSG	540
ATCMWCNCCG	TACCTKTTAA	AATTCTTTGT	GGTGGAACCC	AWYCKAAAAA	NMINIYCCCN	600
TCCAMMGGGG	CYCGGAAKKT	CNACNTGGKT	NACCCCTNCC	YTTGAASTTT	TCYTGNCCCC	660
GGCCCKAAAS	ANACCSGAKC	CCCGGAAYCS	WTAGGCYTCN	TGCCCCSTTA	AATTKGNCYC	720
AATCCKCCAA	CGCTCCCCGG	GGTCSSCCMT	TAAAMTTCCC	CCCKSCASNG	GAATYCYKSG	780
GCWGTMATTW	CCNCCCNTTT	CYYGKNAAAC	SCCCCCWKGN	GSCTYCCCCN	SNTTSSGCCS	840
GGTTSGAMYC	AAAAWTNGGG	MMCNRAGNCG	SGNAMCCSCN	GKKGGGSATW	TKAAYYCYGG	900
GGGGGTCNYC	CCCCRCSNAA	AAGYGTKGGC	KCCSSSCCYC	CCMARTTTYT	CNGGMRCMAM	960
ACCANGGGNG	CTCCCGTNCW	WGGCTCCCSN	SNSMAMAAAN	NKCKCCKGGS	CKGARRNMNA	1020
MCTCSNGNGG	WICCCKNKIC	NSCNSGNCGS	YGGNSASWCC	YNYCNCCACA	ANC	1073

- (2) INFORMATION FOR SEQ ID NO:328:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1166 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:328:

CGCCCCGTTC	TTMMMTTCAY	TCATTCACCG	GGMTCTAGTG	CGGCCGCAAK	CTTGTCKACA	60
GATCTCGAAY	TCGGCAMGAS	ACAATSTCGG	GTKGGGCAAT	GTCNGGTGGG	GCAACTTTGG	120
GCTCGGRAAT	YCGGGGTTAA	CGCCGGGTCT	RATGGGTSTG	GGTAATATCG	GGTTTGGTAA	180
TGCCGGCAGC	TACAATTTCG	GTTTGGCAAA	ATATGGGTGT	GGGCAATATN	GGGTYCGCTA	240
ACACCGSCAS	TGGRAATTYC	GGTATTSGGT	NACCGGTRAY	AAYCTGACCG	GGTNCGGTGG	300
TTYCAATACC	GGTAACGGGA	ATGTSGGTTS	YYYACYCCGS	GSAACGGNWW	YTTNGKTCCT	360
TMMCNCTSSM	CCKSAAMTSM	KMGGTSTYCT	MTYCNNGGAS	TAMTYNMCCC	CCGWAYCKSC	420
WAYCCCTCGT	CATYCCMCMC	SGSGYCCTCA	MNCCACCYTG	NGYYCCCTCC	MKMTCYCAYT	480
CMNTCCGGTW	CCTNTMMNCC	CSCNCRYCTC	AMCNCTKSGK	CACCNATMYC	CSACKCHTCT	540
MCYMCSCAKN	MITCCCCTCN	CCTYTNNCCA	MCMCSCTCTM	TCMAACTCKC	CCGGYCKCNC	600
MYCTCTCKCC	AYNMAACCKK	TYCYWCNWYC	YMYCKCKCAG	WYKNMCICCW	ACTCTMYNTT	660
TCTCTCNKCC	CMKACCKNTT	CTCWCSCCCC	CCACAKAYMC	YAWCMTMTCC	MCTCKACSCC	720
CYYCNNYCCM	NMCWCMTCWC	TWNAKCANCN	TTCTTCTCTC	MMYMTMACKC	WCNNTCNCCK	780
SGACCYTCTC	ACTIONICCION	TCTCCTTMCK	CCYMWCNTCC	MKYNCCCTCC	NMTCMTCKYT	840
CCTCNCNMRY	CYYYAKCAKC	NMCTCCCCAN	KMCAKCTKCT	CCCCCAKMKS	ACNCKCCCWC	900
CCTCCTATCC	WCTCTCWCTY	ATCTCKCTCW	CNYCMYMKMC	ACNCKCYAYT	CNACTMNMWN	960
CCANCNCTCT	CINYCICWCK	ACGTYCKCCK	CTMCKCNYMC	NRWCIYRCCI	CKKCCNCCRN	1020
CKNMCMKCIM	CTCTCCWMKM	TCCCWCCCAT	CTMMKSTCTC	WCNCMTCCCT	CNKCCYNYNT	1080
KCYTYCCMYG	CTTCKNTCMT	MCCWCCYATC	TCTMKCCTCT	CWCACYMCAC	WMTTACWNCC	1140
ACTCTCTRCW	CKCCKCMCCR	MTCTCB				1166

#### (2) INFORMATION FOR SEQ ID NO:329:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1230 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

#### (ii) MOLECULE TYPE: Genomic DNA

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:329:

NGNGGNNNNT	CWTACATCWN	TCTNCACCSG	NGMTCWATTG	CGCGCCGCAW	NCTTGTMNAS	60
AGAATCTCNN	AAYTCGGCAC	ANATGTCTTT	TSTMTAKTGT	GGCGGGGNGC	CACGCCKTAT	120
GTGYGCCTGG	GYTRACCCAA	CCCCGCGGCS	CGGGCCRACC	AGGCGGGGRA	TSCAGGCCGC	180
GGCGGCCGCG	GCGGYTATAT	RAAGCGCCGY	TTTTKTRATA	ACGGTSCCGC	CGCCGGGTRA	240
TTACGGGCAA	AAYCGGKKTT	TTGGGTRTAT	AACGCTAATT	GCAACCAWTT	TTTYCGGGTC	300
AAAAACYCGG	CGWGCANATC	NCGGGYCNCT	RAGGCGCATT	YMCGCCAAAA	WTNTGGGCGC	360
AAAACCCCKT	TSYTATTTTN	TGGGCTATSC	GGYTGCTTCG	GCAAACGCTY	CCCGGGTTAA	420
TCCCKTCCGC	GGCGCCGCCN	AAAAACCACC	AATYCCGYTG	GGGGTGKYCC	CMCAGGCSGT	480
TGCTYCGNGY	CACCTGGCCA	$\lambda \Delta YYCCCAWT$	AKATTGGGTG	SCYCKTSCGG	TTSYTGGGCY	540
CAATTACCCC	CNCGGGNAAA	GRRAAAANAA	ATCNTCCNTT	TGCTCGGYCA	YCTTTMTTGG	600
SAAAAGGGGC	ATGGCSCGGT	TYYTTTACCT	CAAYCCCCNA	NCANTWACCT	YTCCSCCCGG	660
GGGGNCANAA	CGSTINGCIC	CGSGGNAKCC	TKGTMCCCGN	ATCNAAAGGC	CNGAATTTGG	720
TYYSSTYCNA	${\bf ATTWTWKKKY}$	CCCCWCNTTG	YAAAAAKCCA	AAASAKCCCK	YCNCAMMYKT	780
NGGGGTYSSG	GCCKNYCTTK	SNMTTAAACC	CYCCCCAAAA	YYNSGGGKKT	TCCGCYNSAT	840
KCCACCNCCK	GNGGGGGGNA	SAAAAAAAY	TTTYCCSAAA	ATCCCACCYY	TCYKTKSTRY	900
AMACCCCCTT	TYYMKKAYTC	CKYSCNATTC	SGMTTCWAAA	TYCCGYGGCT	TNTTCCCCCK	960
CSGGNGCCCC	AAWTTTGKTT	YNCNANTTYC	CCCNAAMNCM	AWTMGGGGKS	KCCATTCTGG	1020
SCYTMAANTA	AAANAANGGG	NKTTTYYCTY	MANAAACACN	GTGKCNCNCN	CNAAMAAASN	1080
AKMAAAKAGN	KKKMTKNNSA	AANCCNCCCC	CTSTYTNYTT	NKTNMNCKCC	CYGGKKNKGM	1140
SWSWYNTTCT	NCCCRCCCCC	YNYNKTGANA	AAMMNCYCCS	GGSTMCRNAN	ASNMNTTTCK	1200
STSTNGMGCC	KMBASNANAN	MCAMWKWYCC				1230

#### (2) INFORMATION FOR SEQ ID NO:330:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1022 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:330:

MCMCCCIONIN	<b>7</b> 3.1.0.					
NGNGGGKNNA		ACSSGGTCTA	TGCGGCGCAW	CTMGTMAASA	GATCTCNAAY	60
TCGGCAMNAN	GCATMTCMMC	CATATATAAC	CATTGCGTCS	GYWTGCAWCT	CRAAWCTGTC	120
CTTCSKGCCG	TTKTACRAAG	GTGGMWTGYT	CWTYCCTRAA	SCCCTCRATC	TCKTKTATYC	180
CTKGGGCTYC	ACTTTAACSG	RATKSCTGCC	TTKTAYCATT	RATGCAAWTA	WTGGYCRAWT	240
KTTGCAGGCC	RACGGCWYCT	TTTYCCGCRA	GRACAATNGA	TTGGAWYCGC	TYCGCRAGGC	300
CCGGCACCAR	ACCGGGCNCC	AAAGGYCCGC	GCAAWTSCCT	GGKTCAAAAA	TGGTGCAAAC	360
AAAMCNATCC	CCGGYTTRAC	CGCAGYTAMC	ACAAKAAAAT	TCCCWTGGCC	GCACCAWNNT	420
TTYCRATCWY	CWYCCCCACC	TTRAACTTGK	YTGCSGTATT	GCCTKCCTGC	CTCRACAGCM	480
YCNCCCKTCA	AACCTGCGGT	GACTCCAACT	GGTCTGGYCG	AASGGGGGYT	CAMCGGACAA	540
AACCCCRANN	TCGCCAAATT	TTCNCCCCCC	CYCGGGAAAN	GKTGATMTTC	TCSNAACCSA	600
CMGGGNNYTW	NAACCCTGAA	CSSSGSNKGA	MYNSCCSGGA	ANTITTCCCT	TYNGGGCGRN	660
AAANCCTTTT	AAGGTACCCC	KGGNGGGGKG	CCCYYTTGGG	AAAACAACCC	CKATTGGKTT	720
TGGAAATNTT	TKCNCCCCCA			CCMMCTTTTN	TCMSCNMTYY	780
YCYYGGGAAT	TNYTCGCCSG	GAAYYCGGSM	CCKGYCCTAA	NCCCCMNWGG	GKYSTGSNAR	840
GGRATMAWWT	TYSTTTYYMC	CCGGCNNCCC		KGNTGAACMA		900
GSCNMYMWYY	YCNNNGNRTT	TNRGGSSNMT	TYMAAAMMAN	GGGGKYWTYY	CKCCNGSCNN	960
GKTYSGGGST	TTTCCNTTTS	GGGSSATYKG	MACCCCKTMT	AYCCGGGGGT	NTKTKYCCCC	1020
SC						
						1022

#### (2) INFORMATION FOR SEQ ID NO:331:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1083 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:331:

ADATO CONDITIONED A	M31/31/077/07					
MACGMAKATA	TAMAYCWYCT	NCACCSGGGA	TCWATTGCGG	CCGCAATCTT	STMAASAGAT	60
CTCKAAYTCG	GCAMGANCCG	CAWCTATTTG	KGTGRASCGC	ACCAGCGRGA	CCTCGCSGKT	120
CKTTYCTIGC	AGRGAGGCCK	TGGGTGGCRC	CGGTGGCAAT	GCCAACCGCC	CCCCAAAACN	180
CCGCAAATMY	CRAAAAAACAA	CCCSGGGGTA	GKTCCSGGCC	GCCAAATMAA	TAACCGTKTT	240
AACKCAGGCN	ACGGCCAACC	GGYCCCGCCC	AACCAAGCNA	CCTCCCCSCC	NATAGGYCCG	300
GTGGGGGCTG	CCKTATYKCC	AASTCGTCAY	CTCNACGGGM	CGGYCCMCWT	TCCGCCTCAT	360
CCGTCTCTCC	TIMMATTITC	CRTCCACYKG	GCGGGGAACY	TTTTTNYCNC	CCTTGSCMAN	420
CACCNAAGGY	CNAAAATTNC	CCMTGCCKYG	SNNCAAAYGR	GATTGGGGTY	CGKKTTTTNT	480
TCNMCCMAAC	CCCCNTTTNA	CGCCCCMATC	CCYTWATACC	CCCWWMCMNS	ANGKTTGNSA	540
AAKTNNCCCC	AAATRCCAAA	MTTCTTCGCC	NTTIMIWMCY	YYCCTTTCCC	CMCCCWNAAA	600
GGSCCRCCYY	TCGGGAANTY	TCCCCNCAAA	AWTCAMWCCM	TTTCCCNCCA	AGAAWTTCSG	660

RGGGKAAMCC	AGKNTNNTCC	YYYYCCCCAA	YCKTNGGGSK NNTYCCYKGG KAAAATTTCC	RMCYNNYYCY	TTAAANRASR	720 780 840
GCCCCCSAAM YCSCMATAMA	ACTMTWAYTT CTTNGGKCCT	TCCCKCGNNN NTCNYGSGCG	TTTSYCCKCS CMAAANAAGG	KCAMWMWMTG CGCGSTTCTN	KKNCTTTTT TTCWMAMACA	900 960
			TKNNCCCNCC			1020 1080 1083

#### (2) INFORMATION FOR SEQ ID NO:332:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1069 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:332:

GGGGNNKYAT	MCAYCWTCTS	YACSGGGMNC	TATTGCGGCC	GCAWYTNGTM	GASAGATCTC	60
GAAYTCGGCA	MGAAAAAGW	GATGTGCTGG	ACCTTMCCGC	GCGGGACGCR	ACCRACAAAG	120
RAASCGCGCC	ANAATATTGG	CCACAKTTGG	TCACATATTT	ACCCAATTMT	AYCAGGGAYT	180
MCCATTCCKG	GGACCRACCG	CACAATCCCR	ATSKTGGTTT	GCRAACCCTR	ACCGTCCCCA	240
MYTYCGCCRA	STTGAACCAG	GGCRAAAAA	CGGCCRAAWY	CTCGCCCTGA	NTCCCGCTCS	300
GCGCNAATAA	CTAGGCCCAT	TKAACGGAAC	CGGNGGCCSC	NANTTGGCCA	ACAGGTCCTR	360
ACAAAGGGGC	CCCASYYCGG	CCGGWTCCCW	TTYCACNCCC	TNKTCTCKTG	CCGAATYCGG	420
WTCCRATNYC	CCWTGGGCCT	TKTCKYCKYC	KYCGGTNCCA	AWTCTNGGTA	TNCTATRGKG	480
TCCCCTAAAT	SCANATCTGG	GCKYCCATIT	NCTGGSNTTC	NATTTAMMAN	SRRCGGTTCT	540
TTCWTTCCRA	AACCGSNTGG	GCCCNNMCCA	AAAAATGATN	ATAATAATGK	YGSCTTTCAA	600
ACCCCGCCCC	CCCATTCRWT	CSGTTCCANC	CCCCNGNGGT	TAAGKTGGGA	ATTTYTNAMC	660
YCNARGCCCT	NATTTSGGNA	AAAACCYCYC	GGGYCTCAAA	CMNYTTTTTT	GSKSSNTCGG	720
GCTCRTTCSC	CAAAACCCAA	ATTNTYNYGG	GGYCCKTNAA	ACMCGGYCRC	RCCGGAAATT	780
TTTYTGGTTC	AACCCCAACC	TTTTCAASCC	NTTTTYTYYT	TRCCSSCSMN	TNGSSGGGNT	840
KSSCCNTTCY	RARKKCCNMN	GGGGGWYCYN	CCCCRMNTTT	CTTTTTTTT	CCGTNNMAAM	900
NGKTTCTTCA	AASMCCCCCC	SCCCCCNSAA	ACCCCCTNAR	GTTTTYCMMA	AANNWYNNGN	960
KNCCCCCCCC	MMNAAAAAAY	YCSCCCGNRN	ACSMSNGGGA	MCCCCCGGSN	NTTRKTTTTT	1020
TNCMSGYCCC	CSRMASYYTT	TKAMAMANRR	GAMNSMTTTY	TNNRGNWNK		1069

#### (2) INFORMATION FOR SEQ ID NO:333:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1210 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:333:

NGNGGGGKWK	MATACATCWT	TCTTCACGSG	GGATCWATTG	CGGGCCGCAW	TCTNGTMCAA	60
SAGATCTCGA	TYTCGGGCAM	NACCCACCWC	TCCRAAAAA	ACCCRAAWCT	CGGGSKCTYC	120
GARAAGTGTT	GCCCGCKTTR	AATTTAACAA	ATTCAGTGTC	ANAGTGTCAC	GGCKTTACWT	180

YCCCGGCAAA	GGGGCCACAA	CCTGCAGRGA	SCACYCRATG	GKTGYTGKTS	CNCGGGCGGG	240.
CCGGKTNAAG			TMCAAANATC		YCGCTGGRAT	300
MCNCAGGGGT	GTCAAAAAAC	CGCAAACAGG	CACSCCANCC	NTTTACGGGS	CTTAAAANGA	360
AAAAGGGCTG	ATGCCCCCAA	GGGGCCCGC	NCCCAACCTT	CCGTTGGTCA	ACAACCCGGT	420
CTCTCKTGCC	RAATCCGRWT	CCRATNYCNC	CWTGGCCTTK	TCKYCTYCTY	CGGTACCCAA	480
ATCTGGGTAT	CCTATASTGT	CCCCTAAWTT	CCAAATCTGG	GCTGTCCATT	TSCTTGGCNT	540
TCCAAATTTA	CCANCAACGG	TITCTINCAT	NCCAAAAACC		NRACCCRAAA	600
AAATGAATAA	TAATAANNGG		ACCNCCCCC			660
NMNCCCCCAG	NGGKTAGGTK		TCMACCYYCA		TITTNGRAAT	720
KAAACCCTYC	YCNGGGTCWW				MWNCKRKNST	780
SCCAAAATCC	MAAATANTTT	YYTGGTYCNA			GGAAAAWTTTT	840
TINTGKTTSA	ACCCCAAAAC	YTTTTCMNAA	NCSSKTTTTY		AMNWTGGGYS	900
GGGNATKGYG	SCYTNTCTTA	TKTKYTYMTW		MKMTCMMCCC	CCMTTTYYCY	960
NYWRTTTTTN	KCCCCKTNMR	NNRAANNGGN			SCCKNCCCNA	1020
AAAAWCCCCN	NNNARAKINI				YNMNNAAAAA	1020
AATMYCCNCC	RASANMCASM			NNNNTMTTNT	TTTTTTCSRA	1140
GAGCKCCSCG	MNNANMKNCK			GGNGMNCKCC	CCNAGAAMWK	1200
CTKSTCCCKS				301101210100	COLUMN	1210
						1210

- (2) INFORMATION FOR SEQ ID NO:334:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1105 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:334:

NGSSSNGNNA	TMCATCWYCT	GYACSGGGMT	CWATTGCGGC	CGCAACTNGT	MAASAGATCT	60
CGAAYTCGGC		ACCGCCGTGT	MTATACACCG		GTKTGCCAAA	120
ACCGAGACGC	GCCGGCCGCG	GGGYTCCAAC	GCKTTACYTR	ACCCGCCAGY	TCAGTGTTRA	180
AACCGGTGYT	RAGGGCCGCA	CCCAACWTAA	ACGCTTTAKC	CAAGRAWYTG	GKTGGCCCGC	240
AGCCACCTGY	TGTGGYTGCC	CTCWYCGGTG	GTAGCGCCGG	TTANCGCCGG	TTGCGCGYTC	300
AMCASCSCGC	CGGTRATCCC	AKCNWTCCCC	CGGCCMRACC	CACCGGGCAC	TTTGRACGGT	360
GCCGCCAATT	CAAAYCKYCT	GRWTCCTTCM	AAACACCACR	AAGGCCACCM	CCMSCACCNA	420
ATMGGGRACT	TTAAGGCCCA		NTRAKCNCCT	CCCGGGCRAA	GGTCCSGCAA	480
SCRATCCMAA	AAAAKCKNAT	TTCCCCCAGC	AKCAACCCAA	MMCGSTTTGC	TGCTTCCGGA	540
TTCGAAMCCA	ATTMCWGGKT	NCNWGGGAAA	AACASCNNCC	NWTAKCCMGG	CCCMCGGGCA	600
ATTTCSGRAA	SAACCCCTNY	CCCGGGTTTT	YCCTGCTCMG	GCCCAANACC	CCCGGGAATC	660
AAAAASGGTC	GGNCAAANGG	GCMAAACCCS	SACCOMACTT	WTTCCRCTTN	GGGGGGSCWN	720
CCKNGTTTAA	AWKSCCTCYY	CTSCCCAAAY	TCGGKCMAAA	NNGRKTTGGK	TTNGGCNACC	780
NTTTCCGGKC	CCGGGKGKGK	WGKYCTMNMA	CSTTTNTTTT	SCCCCYKAAA	NYSCCCCCC	840
CGGSSCCCCG	CCCGGGGGGA	NNTTTTTAMA	GKKTYCCCCT	CCCCAMAAAA	ANACCCCNYC	900
CCSGGSCCCT	TTKRWAAAMN	KCTSCCCCNG	GNNGGGGKCM	GGKTTATTMT	NNNCCSCCC	960
TCCGCGSAAA	AAATAKMTTT	SYCCCCCCINC	CTCCKNCKNR	GKAMSMSCGC	TCCCYCTCNC	1020
GCNKNTWAAN	ARSNCCKKNN	CCNCYKCCGS	NSNGKCNWCD	NCCSTSSNCT	NKGCNCKNCN	1080
KAAANAAYNC	NGSMSTSSMN	CNKCC				1105

- (2) INFORMATION FOR SEQ ID NO:335:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 936 base pairs

- (B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:335:

NGSNSNKNNN	TAMAYCWYYC	TSCACSNGGA	ACWANTGCGG	CCRMAWCTNS	TMKASAGATC	60
TMGAAYTCGG	CAAGAGCGGC	AAGAGTGTGT	GCATCTGGTC	ANAGTSTMMA	CRCGGTGCCG	120
CSGGTGKGTR	GASCACMCAT	NTGCGRACAC	CAAACCCKTC	GCGGGYCACC	GGCKTCGCCT	180
GCAAAWYCCT	CCAGGCCACC	TCRAACAAYW	YCTYCTGCAA	CGCARGCCGT	TYCGCGGCCG	240
RATCCTGGKT	CASYYCGCCK	TGCGGTGCCC	AAGKTACTGG	CSCAYCAAAA	CCGCTCCGG	300
RAACRAACKT	AAWTYTGCCG	AATTTCNTTC	CCCTGCGCCT	TCATAAATT	NTTN ACCORC	
CGCAAMCCTY	CGGGCKTCTC	CTCKTGCCRA	ATYCGRWTCC	DATAVCCCCA	TOCCOMITMO	360
KYCTYCKYCS	GTACCCAAAT	CTTGGGTATC	CTATANTKYC	COMPANION	TIGGCCINKIC	420
KTCCATKTSC	TGGSKTCCPA	ATTTAMMACA	VCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CCWAAANRCA	AWTCTGGGCK	480
CCCCBACCSA	177786777777	ATTIANMACA	NCGGTTTCTT	TCWTACCAAA	AACCSNTGGG	540
CCCCAACCAA	AAAAKGATAA	TAATAAKGTG	CWWWCAAAAC	CCCGCCCCCC	RRTTCAAYCG	600
GICCARCACC	CCANGNGGTN	AGGTNGGAAT	TYTMAACCCC	CAGCCCATAA	SNTTNSGNAA	660
AAACCCCCCN	GGGYMYCAAA	AMMCTTTTTG	GGGMTTCSGS	CCATKGYKCC	AAAACCAAAA	720
IMITICYGGT	CRWAAAAACC	GGCCCNCCCG	NAAATTTTTT	GKCAACCCCA		
CCNNNTTCYY						780
SNCCCCNAAN	YYCCNAANKG	WCCCCCONG	NGSSCNIIII	TWTTTYYNNA		840
NAAAVACCCC	TTCCMMMM/	NACCCGSNMA		YCMKAAAAAC	CCCCNCNCCC	900
NAAAYACCCC	MAAAKWITCM	AAASMSCNNG	YCCCCC			936

- (2) INFORMATION FOR SEQ ID NO:336:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1042 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:336:

NNNGNKNNNY		YCTSCACCSG	GGNNWCWATT	GCGGCCRMAW	KCTTGTMAAS	60
AGATCTMNAA	YTCGGCACAG	ASSSGCACAG	ASCCGCGGCG	CTATYCMYCC	GYTGCTCATG	120
CTCAACACGC	TCKTCGGCGW	GRATAATGGC	NCGCCGCCGG	CGCCAACACG	YTCAAYTGCT	180
TCGCCAACGC	CATATNTCAA		AAASCAAAAC			240
SCGGRAASCG	GTGCCAACCC					300
TCKTCCTCCT	GGGCTATGGT	GCGCCACAAA	CCTSYTGGCG	WGGGTCTGGC	CCTGGGYCAC	360
CGYCRCMTTT	TATNINICCK	YCTACACNCT			CACMAAATTG	
TTTTGGGKTG	GGGSSGCCGG		TAATAATCSG			420
CCATANCCTG	GCCGGCSCTG	GCAAATTTCC			MYCACCGGWA	480
CTNSAAATCC	GRATCAATNO				CCCCACAMRC	540
RKTNCCCYAA		CCCNKGGCTT		GTRCCCAATY	TGGTTTCTAT	<b>6</b> 00
			YGSTTCCAAN	TTNACAAMAS	GGTTTYTCMT	660
ACCAAAACCC	NTGGSCCNNA	CMNAAAAKNA	RAAAANAKGG	KCTTTYAAAC	CCCCCCTAT	720
TCAWYCGGTN	CMRNWCCCCG	NGKAAGGKGN	GAAAYTTHRA	CCCAANCCMT	ARSTTSGNAK	780
AAACCCYYCG	GGGTSMCAAA	MKNTWTTSSC	CTTCGGMCTT	YCCAAATMSA	AAATYYTCKK	840
KRMNAAAAMC	YGNCCCCSAA	ANATTTTTGT	NAAMCCCKMA	YYTRTTWMCC	WITTICCYCC	900
CCMCNNSNSG	GNTNCCCTTY	TYATTTCYMM	MCRNNSGACN	CCCCMNTYTT	TWTTCKCWCN	960
MMARGSNNYT	RGRMMNMNCC	CCNCCCCNAK			NNKYCKCCCC	1020
CCCMWMINKINC	CCCCMNCMTT	TM			*******	
						1042

#### (2) INFORMATION FOR SEQ ID NO:337:

#### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1073 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:337:

NNSGSGMKKK	ATAMATCWCT	CTSYACCSNG	GMTCWATTGC	GGCCGMAWTC	TNGTMAASAG	60
ATCTCGAAYT	CGGCAAANAK	ACGCMAYGTC	AAGTGTRAYY	CGGTCACATA	TCMTCGCGNG	120
TCAACMCCAA	AGCCGNGTCA	CCGYCTCCCT	GGGGCGCCAC	CCCCATCGGT	RATGCAACYT	180
CGCGCGCCAC	CGYCAAAAGG	KTCWTTRAGG	CGCTAAAGGT	CAMCAATTCC	TRAGGTYMCN	240
CACCGTINIT	TGGCCCGCCC	RAWTYCTRAC	CCGCAATWTC	GGTAATCGGR	AATTTGGGCW	300
YCGGCTTGGG	CAATAAGKTN	TTGGGCAACG	GCGGRWTCYC	NCTGGCCGRA	ATTCCCNCAT	360
TCCKTTAACG	GKTGRACCGT	TTYCCCGGYT	GCCGTAAYTG	YTYCNTGGGC		420
CRNAGCASYY	CRCTAACGGY	CMCCAGGCAA	TACCKTTGGC	TTTRAACCAC	CGGRATNAAY	480
TGKTACCCAC	YTCAASSGTS	CTGRANTTRK	TNTCNTGRAA	AANMCCACCN	AACCCGGNTT	540
RATCTGCTTC	MTCANCWTTT	SCCGGGTTCT		AAYCTTNATC		600
GTTTAMTTTC	CCAANRAATT	CGGYTTGCCA	CCTTGGCCGS	GGCTGGTTTM	CGMWCCTTRR	660
AMATCCNCCS	GCGGGSAAAN	AMTTSGGNTT	SGSCCGGTCC	CCCGNAATAT	YCNTGGNCCT	720
GNAAATTGSS	GGGATCCCCN	GSGNAYCCGG	CCWTKGGGGK	TNCCCAGTTG	GWACAATTYC	780
WKCCGTTCCA	AACCCGGGNC	CGGGGGGTGG	GSCCCNTTTT	CCTMYNNAAA	AAGKGTTTGN	840
NYYTTTTCCG	CNRAANTTCA	CCSKCNKTNT	GGNCCNAACY	YYYCAANTTC	CANACCTTTA	900
AASAAANCYK	YGKTYYCCCC	TTTTMCCSGS	SANCCCCCCM	NMSSKNCGGG		960
TYNGCCTTAN	CNSNKTKTTT	TNKTYCCCCC		NCBKKCNKRY	NGNSNMNCCT	1020
MKYSKCNNNN	SNNNNNKCGN	GSNCSGMKYM	CMNNCNGMYK	NGNKSNNCCC	MSC	1073
					•	

#### (2) INFORMATION FOR SEQ ID NC:338:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1061 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:338:

GNSNGNKNTN	TMCAYCWYCT	SCACSGGGTC	TATTGCGGCC	GCAATYTNGT	CKASAGATCT	60
CGATYTCGGC	AMNANAARTG	TCGTCGTCAA	TTTCAGKKTG	GTCKTCAAAY	GGGCCAGGCC	120
GNGACCRACA	CCCTGNGTCA	CCCAAAANAC	CAACAGCWTC	AAATWTCAAG	GCCRAGGCSC	180
TRTCAATYCC	CRASCAKTTA	ACCGTKTCCW	TCRAAGGTGC	CRAACCAGGC	ACCCAGYTCA	240
CCGCCSGGCA	AWTCGCGCTG	CCGGCCGGTN	TCAGCCTGAT	TYCTGACCCT	RWTCTGTSGG	300
TGGYCAMCNT	GGTGAAGGCC	CWWCCGCCNA	AGAACTGGAG	GGCRAATTCC	CAGGANCONA	360
GRAACCCNAG	GAACCCGCGG	TAKAANCCGG	CRAAACCRAG	GCCGYTGGCN	ATTCCMATTA	420
NAMSGGTTTG	CRACNTGGCC	RAACCGTTTY	CTTGGTCGGC	CTCGGCAACC	CTCCNATIA	480
TACCCCKTNC	CCGGNMCMAC	CYCGGGTNCT	TGKVCCCDAT	MECCYCCCC	CIGGACCANI	
ראנא זייייירייאני	CCCNICCANICT	TTTCCCCCCCC	TORTCCCAAT	MIGCICCCGC	GNRANINGGC	540
CHARITCCAG	GGCMCCAMC I	TTCCGGCCCX	AATTCCCYTG	GTTAATCACC	GGGCNCNCCT	600
GGTTTTGGGC	AACCCCNCYS	$\mathtt{CTTMTTTAAA}$	CATTCCGSCC	CAAATGGGNC	STTGGSAAAT	660

TCTNTYCGGT	GGGGCSGGCR	ANMYTTCTCT	YCCCNAASAN	CTTAMYCCAN	TTCGSSNTCC	720
CGGKCAAAWS	NGGGGGGGNA	AAGGCCCCC	CGGNTSCKCC	GGGGKKGCCC	CYGGKTTCAA	780
AANTTTCSGG	GKTSTMSCGG	NVTCSCCCCC	CSGCCAAGRA	CCGNGGTTTT	TTTTTGAACC	840
KCMANTCSSA	AMCCGCCSSC	CCCMAAAGGS	GCCTNAAWGR	RAYTTNKSCC	CNNAAACSGG	900
CCCCCAKYTY	SGGKTTCNNC	CNCCSGKKGT	CCMTSTTTMM	MRCCCTTTGN	GNKTTTTTAN	960
MGSCCTTNNC	CACCCCCYCK	GGGKCSMNNA	GAAKTMYWKC	CNGGGGNNAN	RSCCCCCCNN	1020
GSGKGGGGKG	MGAGYSCCKT	CTKGCGNCNN	YKNTTTCCCC	C		1061

#### (2) INFORMATION FOR SEQ ID NO:339:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 986 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:339:

GNNGNNNKWN	ATMCAYCWYY	CTSCACCSGG	GMTCWATTGC	GGCCGCAWKY	TNGTMAASAG	60
ATCTMGAAYT	CGGCACANAG	CGGCACAGAG	TGTGTGCATC	TGTGTCANAG	CTGTCAACGC	120
GGTGCCGCSG	GTGGTRASCA	CMCATTGCGR	AACACCAAAC	CCGTCCGCGG	GYCACCGGCX	180
TCGCCTGCAA	AAYCCTCCAG	GCCACCYCRA	AACAAYWYCT	CCTGCAACSC	ARSCCGTTYC	240
GCGGCCGRAT	CCTGGKYCAS	YTCGCCKTGC	GGTGCGCCAA	GGTACTGGCS	CWYCRANACC	300
GCTYCGGGRA	ACCNAACGTA	AATCTTGCCN	AATTTGCNTT	CCCCCTSCCC	TTRATNAATT	360
TGTTAAACCA	CGCAAACCTY	CGGGCKTCTC	CTCKTGCCRA	WTCCGRWTCC	RATNYCGCCA	420
TGGCCTNKTC	KYCTYCKYCS	GTMCCCAAAT	CTTGGTATCC	TATATTGTCC	CTAAATGCAA	480
ATCTKGGCTG	TCCATNTGCT	GGCGTTCAAA	TTWAMANCAG	NGGTTTCTTY	CTTCCNAAAC	540
CCSTTGGCCC	CAAACCNAAA	AATGATNATA	ATAATGGTGC	TNTCAAACCC	CGCNCCCATY	600
CNATCSGKCC	AMMCCCCRGN	GGKTANKKGG	GNAATTCTMM	AACCCCAAGC	CATAASNTTG	660
SGANAAACCY	ИСИСМССУСА	CCAAAACANY	NTINITGGNY	SSNTTCGGMN	YCATGGCTNN	720
CMAAAACCCA	AATACTNYYG	GGYCCAATAA	AAMMMSGGYC	SAMCCGGAAA	WITTIYTIGN	780
KYNAAACCNA	AAKCCTTTTT	CNAACCCDAN	WNTYCCTNCC	RCRCMANTGG	CNSGGARTKT	840
SSSCTTNCCA	$\mathtt{ATGKYCCMAA}$	AGNGGGRANA	CCARCCCCAA	TTCCTNNNTN	KNKNCCCNST	900
TRNAAAAGGG	GKNTYNCMAA	AASCNCCNCC	NCNCTCCCAA	AAKAMCCCCN	AAAGAKNTCN	960
NAANASKYSN	NNNSCCCCCC	CCMMMN				986

- (2) INFORMATION FOR SEQ ID NO:340:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1074 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:340:

NGNGGGNKRN	ATMMAYCWCT	SATYYACCSN	GGMNMWATTG	CGGCCRMAWT	CTNGTMKASA	60
GATCTMGAAA	YTCGGCAAAG	AGYATKCTCG	GGGGCCAGAT	TTNTGGCCCG	CAACCGCCGC	120
ACTTTGCAYW	TCAACAKTCC	SGGTGCCCCA	AAAAAWTCWT	ACCCCCATMC	TYCKTGCASM	180
ASYTGCGCCC	RATTRAACAC	CCGGCCGGCW	TGCTGCGCCA	GGTATTYCAS	CAGYTCAAAY	240
YCTTTKTAGK	TAAAATCCAG	CSGGCGGCCA	CNCAGCCGGG	CGGTKTAGGT	GCCTYCRTCA	300

3 (77)	/2 aa2 aa.						
AIP	MACCAGCY	CGCCCAGGGY	CACCTTGCCC	AAAAYCTCCT	GGGTCAGCCA	AATTYCCGCS	360
CCG	GCCAACM	ACCANCCGCA	TYCTGGCNTC	AATCVCACCC	GGCCCGGTGY	TR R R LOCALDO	
GDA		MANICCCCCAN	77777777	MICICACCG	GGCCCGGIGI	TAAAMMANMA	420
	LICICATO	HAMACCCCCCAN	TCAGCSYTNA	CNGCMACAGC	CCGCCTTCTT	CAMACCGCCA	480
RTA	CCGGGWT	CAACCGGCCS	GTCAAACTCA	ACAGGCGGNC	AGGCCTCCCC	CCCANCAAAC	540
GTC	TTACSCC	MNYAANAAAA	MA A CATTLOTON	THE COCCUS	CASAASNAAA	COGAMISAMAG	540
000		THE PERSON NAMED OF THE PE	MANGRICIGI	TTTCCCCCTC	CASAASNAAA	AANCCCCSGC	600
	GCCTTCN	NMMGGGTTTG	GGGMANANAA	AARCNCCGGN	GGAACGNATC	CGAAAMCTCC	660
CAA	GTCNCMT	TWAWAACYCN	NNAACCCCC	Villaman Cock	AAGGNTCCCC	\	
للمشمل	יידיא כתיציחים	CCCMAGGGGT	<b>77.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.</b>	MUTITIOGGA	AAGGNICCCC	MILIMACCCCC	720
	IMSGKIS	GCGMMITICITY	TAAAAAAATT	CCCCAAAAAG	CCCCGGGAAG	GGTCMAMCTG	780
GGN	IAAATTTC	CAAMCCNWGK	TINTIYNGGT	TMCGGGGGDA	AATTYCNCTC	CCVVABBICCC	
CSS	CSMMNAT	TAVCCMCARM			MI I I CHCIC	CCIIMMMGGG	840
		TATOGRAMMI	TITINNAAWIM	NSGKKTSAMM	YNNKCCMNNN	SNNMSMANNK	900
TNA	MCKCCCN	CCTCNGNGKY	CSCYNCCCSG	GNAGNGGRAS	MKCCNANMAA	AVA CCAMMANA	960
CGG	AAMMCNN	AATKGNNNSC					960
					CNCNKCNSNN	AANRGMRACN	1020
CCC	NSNSGMN	RRGAARMINY	YCCCCCGSKM	GKGNKAAAAW	GKYCCCCCCM	AAAG	1074
							10/4

#### (2) INFORMATION FOR SEQ ID NO:341:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1195 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:341:

NGNGNCNKNT	MTACATCWTT	CTGCACCSGG	GNTCWANTGC	GGCCGCAWKY	TTGTCGASAG	60
ATCTCGAAYT	44414100			NACTCTGGCG	TGTGTACCCC	120
ATTGNGCGCK	TCACGCGCCC	AYTGANCCAK		TGCCGTYCGC	CKTGCGCGGC	180
GGCCTCACGG	CKCTSCWTCT	RAAGGCWTGG	CGCACCGCAT	TCGGTTTTCT	RAACGCTGGG	240
AAAWTGGCCA	GCCGTCTGGC	TCATGGGNTC	TACGCAACGC	CNGCCCCCAA		300
AATCCGGYCC	NTCCTGANCS	CTTTGAAYCC		ACTGGTTGCS	OTTO T T T T T T T T T T	360
TCGAACTTRK	TCNAAATCCC	GCANAKTGTT	TCNTAMGYCC	CNCCGGAAGG		
TTCNGGWANG	TCGGCNKCCG	GCGCTTATCA		ACGGGGAACT		420
KGGGAAAAAG	RRCCTCAATG	MTYGGTCCKC	GCTGCGKANC	CGCSCCCTGK		480
GAAGGCSMAG	GGTTAANGCC	MTTYCNYCCR		SGKWTTYCGG		540
NNKMAMWTTK	TCRGNGGCCW	ATSTSCCGGG		ANACTYCCKW	MGGANKAMNN	600
SAAAGNTKCS	GCGMGTTTTS	SCCKMGANGN	YCTGATTTSA			660
CGAAWKWRKY	CCYAGGGGGM	GNYCSAGCSC	CGMNNATNAG		CCCGGGGTYC	720
TYTNKGGACC	WSCNNCWSAK	ANAACNNKKT		AGNAAGGKTT	RYGSTSKNCC	780
TAAGAGGAGC		CKTGGANGMM	TGCSCCNTMS	AGNKTNKGRT	YCCNKTSTTC	840
TATKSAGMGG		CCSCGTTTKT	GAGWGMGCGC	KYCCCSNKRT	TCNTNGWAAA	900
GGGNTTTGTA	GAGTAKTCGS		TKTGANAAMN	MSMRKNKKTG	CGMGYTCTSC	960
MNNSKTMKMT			WCSGMCMGNG	AGKNKTNNTS	YANTGARCGY	1020
GGCCNCGMNN		GGAGNGCCCC	CSANGMSTGY	NKGGNMSSNG	ARAKGATGGS	1080
GKTCNGSCGC			GGGGGKTGKC	TCKCSCCGNS	CSANGRAGAA	1140
0.11 0.1100 0.00	CGMGGKIGKT	KTKTKNKTGG	YSTCMSSMMM	NAGAAAAGAG	AGGGC	1195

#### (2) INFORMATION FOR SEQ ID NO:342:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3572 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: Genomic DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:342:

CCATCTGATC	GTTGGCAACC	AGCATCGCAG	TGGGAACGAT	GCCCTCATTC	AGCATTTGCA	60
IGGTTTGTTG	AAAACCGGAC	ATGGCACTCC	AGTCGCCTTC	CCGTTCCGCT	ATCGGCTCDA	120
TTTGATTGCG	AGTGAGATAT	' TTATGCCAGC	CAGCCAGACG	CAGACGCGCC	GAGACAGAAC	180
TTAATGGGCC	CGCTAACAGC	GCGATTTGCT	GGTGACCCAA	TGCGACCAGA	TGCTCCACGC	240
CCAGTCGCGT	ACCGTCTTCA	. TGGGAGAAAA	TAATACTGTT	GATGGGTGTC	TGGTCAGAGA	300
CATCAAGAAA	TAACGCCGGA	ACATTAGTGC	AGGCAGCTTC	CACAGCAATG	GCATCCTCCT	360
CATCCAGCGG	ATAGTTAATG	ATCAGCCCAC	TGACGCGTTG	CGCGAGAAGA	TTGTGCACCG	420
CCGCTTTACA	. GGCTTCGACG	CCGCTTCGTT	CTACCATCGA	CACCACCACG	CTGGCACCCA	480
GTTGATCGGC	GCGAGATTTA	ATCGCCGCGA	CAATTTGCGA	CGGCGCGTGC	AGGGCCAGAC	540
TGGAGGTGGC	AACGCCAATC	AGCAACGACT	GTTTGCCCGC	CAGTTGTTGT	GCCACGCGGT	600
TGGGAATGTA	ATTCAGCTCC	GCCATCGCCG	CTTCCACTTT	TTCCCGCGTT	TTCCCACAAA	660
CGTGGCTGGC	CTGGTTCACC	ACGCGGGAAA	CGGTCTGATA	AGAGACACCG	GC3T3 CTCTC	720
CGACATCGTA	TAACGTTACT	GGTTTCACAT	TCACCACCCT	GAATTGACTC	TCTTCCCCC	780
GCLATUATGC	CATACCGCGA	AAGGTTTTGC	GCCATTCGAT	GGTGTCCGGG	ATCTCALCA	840
TUTUUUTTAT	GCGACTCCTG	CATTAGGAAG	CAGCCCAGTA	GTAGGTTGAG	GCCGTTGAGG	900
ACCGCCGCCG	CAAGGAATGG	TGCATGCAAG	GAGATGGCGC	CCAACAGTCC	CCCGGCCACC	960
GGGCCTGCCA	CCATACCCAC	GCCGAAACAA	GCGCTCATGA	GCCCGAAGTG	GCGAGCCCGA	1020
ICIICCCCAT	CGGTGATGTC	GGCGATATAG	GCGCCAGCAA	CCGCACCTGT	GGCGCCGGTG	1080
ATGCCGGCCA	CGATGCGTCC	GGCGTAGAGG	ATCGAGATCT	CGATCCCGCG	AAATTAATAC	1140
GACTCACTAT	AGGGGAATTG	TGAGCGGATA	ACAATTCCCC	TCTAGAAATA	لاستسانتاناسلسل Σ	1200
ACTITAAGAA	GGAGATATAC	ATATGGGCCA	TCATCATCAT	CATCACGTGA	TOGACATOAT	1260
CGGGACCAGC	CCCACATCCT	GGGAACAGGC	GGCGGCGGAG	GCGGTCCAGC	GGGCGCGGGA	1320
TAGCG'. CGAT	GACATCCGCG	TCGCTCGGGT	CATTGAGCAG	GACATGGCCG	TGGACAGCGC	1380
CARCHAGATC	ACCTACCGCA	TCAAGCTCGA	AGTGTCGTTC	AAGATGAGGC	CGGCGCZACC	1440
GAGGGGCTCG	AAACCACCGA	GCGGTTCGCC	TGAAACGGGC	GCCGGCGCCG	GTACTGTCGC	1500
GACTACCCCC	GCGTCGTCGC	CGGTGACGTT	GGCGGAGACC	GGTAGCACGC	TCCTCTACC	1560
GCTGTTCAAC	CTGTGGGGTC	CGGCCTTTCA	CGAGAGGTAT	CCGAACGTCA	CGATCACCGC	1620
LAGGGGALC	GGTTCTGGTG	CCGGGATCGC	GCAGGCCGCC	GCCGGGACGG	TCAACATTGG	1680
GGCCTCCGAC	GCCTATCTGT	CGGAAGGTGA	TATGGCCGCG	CACAAGGGGC	TGATGAACAT	1740
COCCCTAGCC	ATCTCCGCTC	AGCAGGTCAA	CTACAACCTG	CCCGGAGTGA	GCGAGCACCT	1800
CAAGCTGAAC	GGAAAAGTCC	TGGCGGCCAT	GTACCAGGGC	ACCATCAAAA	CCTGGGACGA	1860
CCCGCAGATC	GCTGCGCTCA	ACCCCGGCGT	GAACCTGCCC	GGCACCGCGG	TAGTTCCGCT	1920
GCACLGCTCC	GACGGGTCCG	GTGACACCTT	CTTGTTCACC	CAGTACCTGT	CCAAGCAAGA	1980
TCCCGAGGGC	TGGGGCAAGT	CGCCCGGCTT	CGGCACCACC	GTCGACTTCC	CGGCGGTGCC	2040
366166616	GGTGAGAACG	GCAACGGCGG	CATGGTGACC	GGTTGCGCCG	AGACACCGGG	2100
CIGCGIGGCC	TATATCGGCA	TCAGCTTCCT	CGACCAGGCC	AGTCAACGGG	GACTCGGCGA	2160
GGCCCAACTA	GGCAATAGCT	CTGGCAATTT	CTTGTTGCCC	GACGCGCAAA	GCATTCAGGC	2220
CGCGGCGCT	GGCTTCGCAT	CGAAAACCCC	GGCGAACCAG	GCGATTTCGA	TGATCGACGG	2280
GUUUGUUUG	GACGGCTACC	CGATCATCAA	CTACGAGTAC	GCCATCGTCA	ACAACCGGCA	2340
AAAGGACGCC	GCCACCGCGC	AGACCTTGCA	GGCATTTCTG	CACTGGGCGA	TCACCGACGG	2400
UAACAAGGCC	TCGTTCCTCG	ACCAGGTTCA	TTTCCAGCCG	CTGCCGCCCG	CGGTGGTGAA	2460
GILGICIGAL	GCGIIGAICG	CGACGATTTC	CAGCGCTGAG	ATGAAGACCG	ATGCCGCTAC	2520
CCTCGCGCAG	GAGGCAGGTA	ATTTCGAGCG	GATCTCCGGC	GACCTGAAAA	CCCAGATCGA	2580
CCAGGTGGAG	TCGACGGCAG	GTTCGTTGCA	GGGCCAGTGG	CGCGGCGCGCG	CGGGGACGGC	2640
CGCCCAGGCC	GCGGTGGTGC	GCTTCCAAGA	AGCAGCCAAT	AAGCAGAAGC	AGGAACTCGA	2700
COMUNICIOS	ACGAATATTC	GTCAGGCCGG	CGTCCAATAC	TCGAGGGCCG	ACGAGGAGCA	2760
GCAGCAGGCG	CIGICCICGC	AAATGGGCTT	TGGATTCAGC	TTCGCGCTGC	CTGCTCCCTC	2020
GGTGGAGTCT	GACGCCGCCC	ACTTCGACTA	CGGTTCAGCA	CTCCTCAGCA	AAACCACCCC	2880
ACCCCCCA	TTTCCCGGAC	AGCCGCCGCC	GGTGGCCAAT	GACACCCGTA	TCGTGCTCGG	2940
CCGGCLAGAC	CAAAAGCTTT	ACGCCAGCGC	CGAAGCCACC	GACTCCAAGG	CCGCGGGGGG	3000
GTTGGGCTCG	GACATGGGTG	AGTTCTATAT	GCCCTACCCG	GGCACCCGGA	TCAACCAGGA	3060
						3000

268

CAGCGATCCG GAACCCGGTG CCCGCCGCCG GACCTTACCG CCACCACCAC	AGTAAGCCGA GACGCCGGGC GACAAGGGCG GCGCCGGCCG	ACGGCCAGAT CCCCTCAGCG CGGCCAAGGC GGGAAGTCGC TCTGCAGATA CGGCTGCTAA	CTGGACGGGC CTGGTTTGTG GCTGGCCGAA TCCTACCCCG TCCATCACAC CAAAGCCCGA	TCGTATTACG GTAATCGGCT GTATGGCTCG TCGATCCGGC ACGACACCGA TGGCGGCCGC AAGGAAGCTG	CGCCCGCGGC GGACCGCCAA CTTTGGTCGC CACCGCAGCG TCGAGCACCA AGTTGGCTGC	3120 3180 3240 3300 3360 3420 3480
TGCCACCGCT	GAGCAATAAC AAAGGAGGAA	TAGCATAACC	CCTTGGGGCC	TCTAAACGGG	TCTTGAGGGG	3540 3572

- (2) INFORMATION FOR SEQ ID NO:343:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:343:

Val Gln Phe Gln Ser Gly Gly Asp Asn Ser Pro Ala Val Tyr Xaa Xaa 1 5 10 15 Asp Gly Xaa Arg

- (2) INFORMATION FOR SEQ ID NO:344:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:344:

Thr Thr Val Pro Xaa Val Thr Glu Ala Arg
1 5 10

- (2) INFORMATION FOR SEQ ID NO:345:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:345:

Thr Thr Pro Ser Xaa Val Ala Phe Ala Arg

27

269

1 5 10

- (2) INFORMATION FOR SEQ ID NO:346:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 12 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:346:

Asp Ala Gly Lys Xaa Ala Gly Xaa Asp Val Xaa Arg

- (2) INFORMATION FOR SEQ ID NO:347:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:347:

Thr Xaa Glu Glu Xaa Gln Glu Ser Phe Asn Ser Ala Ala Pro Gly Asn 1 5 10 15 Xaa Lys

- (2) INFORMATION FOR SEQ ID NO:348:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 27 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Other
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:348:

CTAGTTAGTA CTCAGTCGCA GACCGTG

(2) INFORMATION FOR SEQ ID NO:349:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

WO 99/42076 PCT/US99/03268

270

- (ii) MOLECULE TYPE: Other
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:349:

#### GCAGTGACGA ATTCACTTCG ACTCC

25

- (2) INFORMATION FOR SEQ ID NO:350:
- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2412 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:350:

```
CATATGGGCC ATCATCATCA TCATCACGTG ATCGACATCA TCGGGACCAG CCCCACATCC
 60
TGGGAACAGG CGGCGGCGGA GGCGGTCCAG CGGGCGCGGG ATAGCGTCGA TGACATCCGC
 120
GTCGCTCGGG TCATTGAGCA GGACATGGCC GTGGACAGCG CCGGCAAGAT CACCTACCGC
 180
ATCAAGCTCG AAGTGTCGTT CAAGATGAGG CCGGCGCAAC CGAGGGGGCTC GAAACCACCG
 240
AGCGGTTCGC CTGAAACGGG CGCCGGCGCC GGTACTGTCG CGACTACCCC CGCGTCGTCG
CCGGTGACGT TGGCGGAGAC CGGTAGCACG CTGCTCTACC CGCTGTTCAA CCTGTGGGGT 360
CCGGCCTTTC ACGAGAGGTA TCCGAACGTC ACGATCACCG CTCAGGGCAC CGGTTCTGGT
 420
GCCGGGATCG CGCAGGCCGC CGCCGGGACG GTCAACATTG GGGCCTCCGA CGCCTATCTG
 480
TCGGAAGGTG ATATGGCCGC GCACAAGGGG CTGATGAACA TCGCGCTAGC CATCTCCGCT
 540
CAGCAGGTCA ACTACAACCT GCCCGGAGTG AGCGAGCACC TCAAGCTGAA CGGAAAAGTC
 600
CTGGCGGCCA TGTACCAGGG CACCATCAAA ACCTGGGACG ACCCGCAGAT CGCTGCGCTC
 660
AACCCCGGCG TGAACCTGCC CGGCACCGCG GTAGTTCCGC TGCACCGCTC CGACGGGTCC
 720
GGTGACACCT TCTTGTTCAC CCAGTACCTG TCCAAGCAAG ATCCCGAGGG CTGGGGCAAG
 780
TCGCCCGGCT TCGGCACCAC CGTCGACTTC CCGGCGGTGC CGGGTGCGCT GGGTGAGAAC
GGCAACGGCG GCATGGTGAC CGGTTGCGCC GAGACACCGG GCTGCGTGGC CTATATCGGC
 900
ATCAGCTTCC TCGACCAGGC CAGTCAACGG GGACTCGGCG AGGCCCAACT AGGCAATAGC
TCTGGCAATT TCTTGTTGCC CGACGCGCAA AGCATTCAGG CCGCGGCGGC TGGCTTCGCA 1020
TCGAAAACCC CGGCGAACCA GGCGATTTCG ATGATCGACG GGCCCGCCCC GGACGGCTAC
 1080
CCGATCATCA ACTACGAGTA CGCCATCGTC AACAACCGGC AAAAGGACGC CGCCACCGCG
 1140
CAGACCTTGC AGGCATTTCT GCACTGGGCG ATCACCGACG GCAACAAGGC CTCGTTCCTC
 1200
GACCAGGTTC ATTTCCAGCC GCTGCCGCCC GCGGTGGTGA AGTTGTCTGA CGCGTTGATC 1260
GCGACGATTT CCAGCGCTGA GATGAAGACC GATGCCGCTA CCCTCGCGCA GGAGGCAGGT 1320
AATTTCGAGC GGATCTCCGG CGACCTGAAA ACCCAGATCG ACCAGGTGGA GTCGACGGCA 1380
GGTTCGTTGC AGGGCCAGTG GCGCGGCGCG GCGGGGACGG CCGCCCAGGC CGCGGTGGTG 1440
CGCTTCCAAG AAGCAGCCAA TAAGCAGAAG CAGGAACTCG ACGAGATCTC GACGAATATT 1500
CGTCAGGCCG GCGTCCAATA CTCGAGGGCC GACGAGGAGC AGCAGCAGGC GCTGTCCTCG 1560
CAAATGGGCT TTGTGCCCAC AACGGCCGCC TCGCCGCCGT CGACCGCTGC AGCGCCACCC
 1620
GCACCGGCGA CACCTGTTGC CCCCCCACCA CCGGCCGCCG CCAACACGCC GAATGCCCAG 1680
CCGGGCGATC CCAACGCAGC ACCTCCGCCG GCCGACCCGA ACGCACCGCC GCCACCTGTC
 1740
ATTGCCCCAA ACGCACCCCA ACCTGTCCGG ATCGACAACC CGGTTGGAGG ATTCAGCTTC
 1800
GCGCTGCCTG CTGGCTGGGT GGAGTCTGAC GCCGCCCACT TCGACTACGG TTCAGCACTC
 1860
CTCAGCAAAA CCACCGGGGA CCCGCCATTT CCCGGACAGC CGCCGCCGGT GGCCAATGAC
 1920
ACCCGTATCG TGCTCGGCCG GCTAGACCAA AAGCTTTACG CCAGCGCCGA AGCCACCGAC
TCCAAGGCCG CGGCCCGGTT GGGCTCGGAC ATGGGTGAGT TCTATATGCC CTACCCGGGC
 2040
ACCCGGATCA ACCAGGAAAC CGTCTCGCTC GACGCCAACG GGGTGTCTGG AAGCGCGTCG
 2100
TATTACGAAG TCAAGTTCAG CGATCCGAGT AAGCCGAACG GCCAGATCTG GACGGGCGTA
 2160
```

ATCGGCTCGC	CCGCGGCGAA	CGCACCGGAC	GCCGGGCCCC	CTCAGCGCTG	GTTTGTGGTA	2220
TGGCTCGGGA	CCGCCAACAA	CCCGGTGGAC	AAGGGCGCGG	CCAAGGCGCT	GGCCGAATCG	2280
ATCCGGCCTT	TGGTCGCCCC	GCCGCCGGCG	CCGGCACCGG	CTCCTGCAGA	GCCCGCTCCG	2340
GCGCCGGCGC	CGGCCGGGGA	AGTCGCTCCT	ACCCCGACGA	CACCGACACC	GCAGCGGACC	2400
TTACCGGCCT	GA					2412

#### (2) INFORMATION FOR SEQ ID NO:351:

#### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 802 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

#### (ii) MOLECULE TYPE: protein

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:351:

Met 1	Gly	His	His	His 5	His	His	His	Val	Ile 10	Asp	Ile	Ile	Gly	Thr 15	Ser
Pro	Thr	Ser	Trp 20	Glu	Gln	Ala	Ala	Ala 25	Glu	Ala	Val	Gln	Arg 30	Ala	Arg
Asp	Ser	Val 35	Asp	Asp	Ile	Arg	Val 40	Ala	Arg	Val	Ile	Glu 45	Gln	Asp	Met
Ala	Val 50	Asp	Ser	Ala	Gly	Lys 55	Ile	Thr	Tyr	Arg	Ile 60	Lys	Leu	Glu	Val
65					70					Gly 75					80
-				85	-		_		90	Thr				95	
			100					105		Gly			110		
		115					120			His		125			
Val	Thr 130	Ile	Thr	Ala	Gln	Gly 135	Thr	Gly	Ser	Gly	Ala 140	Gly	Ile	Ala	Gln
Ala 145	Ala	Ala	Gly	Thr	Val 150	Asn	Ile	Gly	Ala	Ser 155	Asp	Ala	Tyr	Leu	Ser 160
Glu	Gly	Asp	Met	Ala 165	Ala	His	ŗĀs	Gly	Leu 170	Met	Asn	Ile	Ala	Leu 175	Ala
Ile	Ser	Ala	Gln 180	Gln	Val	Asn	Tyr	Asn 185		Pro	Gly	Val	Ser 190	Glu	His
Leu	ŗås	Leu 195	Asn	Gly	Lys	Val	Leu 200		Ala	Met	Tyr	Gln 205		Thr	Ile
Lys	Thr 210		Asp	Asp	Pro	Gln 215		Ala	Ala	Leu	Asn 220		Gly	Val	Asn
Leu 225	Pro	Gly	Thr	Ala	Val 230		Pro	Leu	His	Arg 235		Asp	Gly	Ser	Gly 240
Asp	Thr	Phe	Leu	Phe 245		Gln	Tyr	Leu	Ser 250		Glm	Asp	Pro	Glu 255	Gly
Trp	Gly	Lys	Ser 260		Gly	Phe	e Gly	7 Thr 265		Val	Asp	Phe	270		Val
Pro	Gly	Ala 275		. Gly	Glu	Asn	Gl _y 280		ı Gly	gly	Met	Val 289		Gly	Cys
Ala	Glu	Thr	Pro	Gly	. Cha	Val	. Alā	туг	Ile	e Gly	r Ile	Ser	Phe	Leu	Asp

	290					295					300				
Gln	Ala	Ser	Gln	Arg	Gly	Leu	Gly	Glu	Ala	Gln	Leu	Glv	Asn	Ser	Ser
305				_	310		•			315		1			320
Glv	Asn	Phe	Leu	Leu	Pro	Asp	Ala	Gl n	Car		Gln.	A 3 -	x 7 -	N1 -	310
				325		·wp	~~~	GIII		115	G111	ALA	ALA		ALA
Glv	Dhe	777	c		71	B			330			_		335	
Gry	FILE	Ala	Ser	гλа	rnr	Pro	Ala		Gin	Ala	Ile	Ser	Met	Ile	Asp
	_		340					345					350		
GIA	Pro	Ala	Pro	Asp	Gly	Tyr	Pro	Ile	Ile	Asn	Tyr	Glu	Tyr	Ala	Ile
		355					360					365			
Val	Asn	Asn	Arg	Gln	Lys	Asp	Ala	Ala	Thr	Ala	Gln	Thr	Leu	Gln	Ala
	370					375					380				
Phe	Leu	His	Trp	Ala	Ile	Thr	Asp	Glv	Agn	Larg	Ala	Sar	Dhe	T.011	7
385			-		390			,		395			- 110	шец	
	Val	Hie	Phe	Gla		Lau	Dec	D==	77.		**- 1	•	-	_	400
		****	- 110		FIU	Den	PIO	PIO		var	Val	ràs	Leu		Asp
77-	T	77.	*1-	405	-1-		_		410					415	
ALG	Leu	TTG	Ala	Thr	TTE	Ser	Ser	Ala	Glu	Met	Lys	Thr	Asp	Ala	Ala
	_		420					425					430		
Thr	Leu	Ala	Gln	Glu	Ala	Gly	Asn	Phe	Glu	Arg	Ile	Ser	Gly	Asp	Leu
		435					440					445			
Lys	Thr	Gln	Ile	Asp	Gln	Val	Glu	Ser	Thr	Ala	Glv	Ser	Leu	Gln	Glv
	450			-		455					460			<b></b>	<b>-</b> -,
Gln	Trp	Arg	Gly	Ala	Ala		Thr	Δla	λla	Gln		777	77-7	77-1	»
465	•	_	2		470	1		n_u	Ala	475	ALA	ALA	Val	val	
	Gln	Glu	712	717		T	<b>71</b> -	<b>*</b>	~1		_	_			480
1116	G111	GIU	Ala	ALA	MSII	Lys	GIU	гÀг		GIU	Leu	Asp	Glu	Ile	Ser
<b></b>			_	485		_			490					495	
Thr	Asn	ile	Arg	Gln	Ala	Gly	Val	Gln	Tyr	Ser	Arg	Ala	Asp	Glu	Glu
			500					505					510		
Gln	Gln	Gln	Ala	Leu	Ser	Ser	Gln	Met	Gly	Phe	Val	Pro	Thr	Thr	Ala
		515					520		•			525			
Ala	Ser	Pro	Pro	Ser	Thr	Ala	Ala	Ala	Pro	Pro	Δla		Δla	Thr	220
	530					535					540			****	110
Val	Ala	Pro	Pro	Pro	Pro		31 a	315	100	mh-		١	31-	<b>71</b> -	<b>3</b>
545					550		A44	ALA	ASII		PIO	ASII	ALA	GIII	
	) cn	Dwo	7.00	315		D	2	_		555	_		_		560
Gry	wab	PIO	Asn		Ala	Pro	220	Pro		yzb	Pro	λsn	Ala	pro	Pro
<b>5</b>				565	_				570					575	
Pro	SLO	Val	Ile	Ala	Pro	Asn	Ala	Pro	Gln	Pro	Val	Arg	Ile	Asp	Asn
			580					585					590		
Pro	Val	Gly	Gly	Phe	Ser	Phe	Ala	Leu	Pro	Ala	Gly	Trp	Val	Glu	Ser
		595					600				-	605			
Asp	Ala	Ala	His	Phe	Asp	Tvr	Glv	Ser	Ala	Leu	T.em		Lve	Thr	Thr
	610				•	615					620	501	шуз		
Gly		Pro	Pro	Phe	Pro		Gl n	Dro	Dro	D==		31-	n	<b>3</b>	mb
625					630	Gry	GIII	210	PIO		Val	АТА	Asn	Asp	
	Tla	175.1	Ŧ	<b>a</b> 1			_		_	635		_			640
Arg	Lie	ART	Leu	GIY	Arg	Leu	Asp	Gln		Leu	Tyr	Ala	Ser	Ala	Glu
				645					650					<b>65</b> 5	
Ala	Thr	Asp	Ser	Lys	Ala	Ala	Ala	Arg	Leu	Gly	Ser	Asp	Met	Gly	Glu
			660					665					670		
Phe	Tyr	Met	Pro	Tyr	Pro	Gly	Thr	Arg	Ile	Asn	Gln	Glu	Thr	Val	Ser
		675				_	68Q	-				685			
Leu	Asp	Ala	Asn	Glv	Val	Ser	Glv	Ser	Δ1 =	Sar	T-1-	Tur	C1	17-1	t
	690			1		695	<u>1</u>	J-C-1	A_ a	261		TAT	GT II	ATT	пĀЗ
		Δας	Dra	Ca-	Tura		λ	~1	<b>~</b> 3 ·		700	<b></b> 1			
Phe 705		بإصد	± 1 0	Jer		FIO	ASII	σтУ	GIN		rrp	Inr	GIY	val	
	C - · ·	D-4 -	83 -	<b>n</b> 7 -	710		_	_		715					720
Gly	ser	PLO	АТА	ALA	Asn	ALA	Pro	Asp		Gly	Pro	Pro	Gln	Arg	Trp
				725					730					735	

Phe Val Val Trp Leu Gly Thr Ala Asn Asn Pro Val Asp Lys Gly Ala 740 745 750	
Ala Lys Ala Leu Ala Glu Ser Ile Arg Pro Leu Val Ala Pro Pro Pro 755 760 765	
Ala Pro Ala Pro Ala Pro Ala Glu Pro Ala Pro Ala Pro Ala Pro Ala 770 780	
Gly Glu Val Ala Pro Thr Pro Thr Pro Thr Pro Gln Arg Thr Leu 785 790 795 800 Pro Ala	
(2) INFORMATION FOR SEQ ID NO:352:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 34 base pairs (B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: Other	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:352:	
GGATCCAAAC CACCGAGCGG TTCGCCTGAA ACGG	34
(2) INFORMATION FOR SEQ ID NO:353:	
<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 37 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
(ii) MOLECULE TYPE: Other	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:353:	
CGCTGCGAAT TCACCTCCGG AGGAAATCGT CGCGATC	37
(2) INFORMATION FOR SEQ ID NO:354:	
<ul> <li>(i) SEQUENCE CHARACTERISTICS:</li> <li>(A) LENGTH: 1962 base pairs</li> <li>(B) TYPE: nucleic acid</li> <li>(C) STRANDEDNESS: single</li> <li>(D) TOPOLOGY: linear</li> </ul>	
(ii) MOLECULE TYPE: cDNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:354:	
CATATGGGCC ATCATCATCA TCATCACGGA TCCAAACCAC CGAGCGGTTC GCCTGAAAC	CG 60
GGCGCCGGCG CCGGTACTGT CGCGACTACC CCCGCGTCGT CGCCGGTGAC GTTGGCGG	AG 120
ACCGGTAGCA CGCTGCTCTA CCCGCTGTTC AACCTGTGGG GTCCGGCCTT TCACGAGACTATCCGAACG TCACGATCAC CGCTCAGGGC ACCGGTTCTG GTGCCGGGAT CGCGCAGGG	GG 180 CC 240

GCCGCCGGGA	CGGTCAACAT	TGGGGCCTCC	GACGCCTATC	TGTCGGAAGG	TGATATGGCC	300
GCGCACAAGG	GGCTGATGAA	CATCGCGCTA	GCCATCTCCG	CTCAGCAGGT	CAACTACAAC	360
CIGCCCGGAG	TGAGCGAGCA	CCTCAAGCTG	AACGGAAAAG	TCCTGGCGGC	CATGTACCAG	420
GGCACCATCA	AAACCTGGGA	CGACCCGCAG	ATCGCTGCGC	TCAACCCCGG	CGTGAACCTG	480
CCCGGCACCG	CGGTAGTTCC	GCTGCACCGC	TCCGACGGGT	CCGGTGACAC		540
ACCCAGTACC		AGATCCCGAG				600
ACCGTCGACT	TCCCGGCGGT	GCCGGGTGCG	CTGGGTGAGA	ACGGCAACGG	CGGCATGGTG	660
ACCGGTTGCG	CCGAGACACC	GGGCTGCGTG	GCCTATATCG	GCATCAGCTT	CCTCGACCAG	720
GCCAGTCAAC	GGGGACTCGG	CGAGGCCCAA	CTAGGCAATA	GCTCTGGCAA	TTTCTTGTTG	780
CCCGACGCGC	AAAGCATTCA	GGCCGCGGCG	GCTGGCTTCG	CATCGAAAAC	CCCGGCGAAC	840
CAGGCGATTT	CGATGATCGA	CGGGCCCGCC	CCGGACGGCT	ACCCGATCAT	CAACTACGAG	900
TACGCCATCG	TCAACAACCG	GCAAAAGGAC	GCCGCCACCG	CGCAGACCTT	GCAGGCATTT	960
CTGCACTGGG	CGATCACCGA	CGGCAACAAG	GCCTCGTTCC	TCGACCAGGT	TCATTTCCAG	1020
CCGCTGCCGC	CCGCGGTGGT	GAAGTTGTCT	GACGCGTTGA	TCGCGACGAT	TTCCTCCGGA	1080
GGTGGCAGTG	GGGGAGGCTC	AGGTGGAGGT	TCTGGCGGGA	GCGTGCCCAC	AACGGCCGCC	1140
TCGCCGCCGT	CGACCGCTGC	AGCGCCACCC	GCACCGGCGA	CACCTGTTGC	CCCCCCACCA	1200
CCGGCCGCCG	CCAACACGCC	GAATGCCCAG	CCGGGCGATC	CCAACGCAGC	ACCTCCGCCG	1260
GCCGACCCGA	ACGCACCGCC	GCCACCTGTC	ATTGCCCCAA	ACGCACCCCA	ACCTGTCCGG	1320
ATCGACAACC	CGGTTGGAGG	ATTCAGCTTC	GCGCTGCCTG	CTGGCTGGGT	GGAGTCTGAC	1380
GCCGCCCACT	TCGACTACGG	TTCAGCACTC	CTCAGCAAAA	CCACCGGGGA	CCCGCCATTT	1440
CCCGGACAGC	CGCCGCCGGT	GGCCAATGAC	ACCCGTATCG	TGCTCGGCCG	GCTAGACCAA	1500
AAGCTTTACG	CCAGCGCCGA	AGCCACCGAC	TCCAAGGCCG	CGGCCCGGTT	GGGCTCGGAC	1560
ATGGGTGAGT	TCTATATGCC	CTACCCGGGC	ACCCGGATCA	ACCAGGAAAC		1620
	GGGTGTCTGG	AAGCGCGTCG	TATTACGAAG	TCAAGTTCAG	CGATCCGAGT	1680
AAGCCGAACG	GCCAGATCTG	GACGGGCGTA	ATCGGCTCGC	CCGCGGCGAA	CGCACCGGAC	1740
GCCGGGCCCC	CTCAGCGCTG	GTTTGTGGTA	TGGCTCGGGA	CCGCCAACAA	CCCGGTGGAC	1800
AAGGGCGCGG	CCAAGGCGCT	GGCCGAATCG		TGGTCGCCCC		1860
CCGGCACCGG	CTCCTGCAGA	GCCCGCTCCG	GCGCCGGCGC	CGGCCGGGGA		1920
ACCCCGACGA	CACCGACACC	GCAGCGGACC	TTACCGGCCT	GA		1962
						1702

## (2) INFORMATION FOR SEQ ID NO:355:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 652 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:355:

Ţ				His 5					10					15	
			20	Ala				25					3.0		
Ser	Pro	Val 35	Thr	Leu	Ala	Glu	Thr 40	Gly	Ser	Thr	Leu	Leu 45	Tyr	Pro	Leu
Phe	Asn 50	Leu	Trp	Gly	Pro	Ala 55	Phe	His	Glu	Arg	Tyr	Pro	Asn	Val	Thr
Ile 65	Thr	Ala	Gln	Gly	Thr 70	Gly	Ser	Gly	Ala	Gly 75	Ile	Ala	Gln	Ala	Ala 80
				Asn 85					90					95	
qzA	Met	Ala	Ala	His	Lys	Gly	Leu	Met	Asn	Ile	Ala	Leu	Ala	Ile	Ser

Δ7:	a (3) r	. 61+	100				_	105	5				110	)	
		11:					120	)				125			
	130	,	/ Lys			135	5				140	١			
Tr:	Asp	Asp	Pro	Gln	Ile 150	Ala	Ala	Leu	. Asn	Pro	Gly	v Val	Asn	Leu	
		Ala	val	Val	Pro		His	Arg			Gly	ser Ser	Gly	Asp	160 Thr
Phe	Leu	Phe	Thr	165 Gln		Leu	Ser	Lys	170 Gln	Asp	Pro	Glu	Gly	175 Trp	Gly
		Pro	Gly					185					190		
		195					200					205			
	210		Glu			215					220				
22,			Cys		230					225	Phe	Leu			240
Ser	Gln	Arg	Gly	Leu 245	Gly	Glu	Ala	Gln	Leu 250	Gly	Asn	Ser	Ser		Asn
Phe	Leu	Leu	Pro 260	Asp	Ala	Gln	Ser	Ile 265	Gln	Ala	Ala	Ala		255 Gly	Phe
Ala	Ser	Lys 275	Thr	Pro	Ala	Asn	Gln	Ala	Ile	Ser	Met		270 Asp	Gly	Pro
Ala	Pro 290	-	Gly	Tyr	Pro	Ile	280 Ile	Asn	Tyr	Glu	Tyr	285 Ala	Ile	Val	Asn
Asn	Arg		Lys		Ala	295					300				
303					310					315					220
			Ile	343					330					775	
			Pro 340					345					350		
		222	Ile				360					365	Ser		
Gly	Ser 370	Gly	Gly	Ser	Val	Pro 375	Thr	Thr	Ala	Ala	Ser 380	Pro	Pro	Ser	Thr
Ala 385	Ala	Ala	Pro	Pro	Ala 390		Ala	Thr	Pro	Val	Ala	Pro	Pro	Pro	Pro
	Ala	Ala	Asn	Thr		) en	11-	~1 <b>~</b>	D	395	_	_	_		400
			Asn	402					410					415	
			Ala 420					425					420		
		433	Gln				440					445			
Phe	Ala 450	Leu	Pro	Ala	Gly	Trp 455	Val	Glu	Ser	Asp	Ala 460	Ala	His	Phe	Asp
Tyr 465	Gly	Ser	Ala	Leu	Leu 470		Lys	Thr	Thr		Asp	Pro	Pro	Phe	
	Gln	Pro	Pro	Pro		Ala	Asn	Asp	Thr	475 Arg	Ile	Val	Leu	Gly	480 Arg
Leu	Asp	Gln	Lys	485 Leu	Tyr	Ala	Ser	Ala	490 Glu	Ala	Thr	Asp	Ser	495 Lys	Ala
			500					505					E10		
		272	Leu				520					525			
GT.7.	7hr	arg	Ile	Asn	Gln	Glu 535	Thr	Val	Ser	Leu	Asp 540	Ala	Asn	Gly	Val

Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys Phe Ser Asp Pro Ser Lys Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly Thr Ala Asn Asn Pro Val Asp Lys Gly Ala Ala Lys Ala Leu Ala Glu Ser Ile Arg Pro Leu Val Ala Pro Pro Pro Ala Pro Ala Pro Ala Pro Ala Glu Pro Ala Pro Ala Pro Ala Gly Glu Val Ala Pro Thr Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu Pro Ala 

#### CLAIMS

- 1. A polypeptide comprising an immunogenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:
  - (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 120)
  - (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser: (SEQ ID No. 121)
  - (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 122)
  - (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 123)
  - (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val; (SEQ ID No. 124)
  - (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro: (SEQ ID No. 125)
  - (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser; (SEQ ID No. 126)
  - (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 127)
  - (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn: (SEQ ID No. 128) and
  - (j) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136)

wherein Xaa may be any amino acid.

2. A polypeptide comprising an immunogenic portion of an M. tuberculosis antigen, or a variant of said antigen that differs only in conservative

substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

- (a) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID No. 129) and
- (b) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No. 137), wherein Xaa may be any amino acid.
- 3. A polypeptide comprising an immunogenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 1, 2, 4-10, 13-25, 52, 99 and 101, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 1, 2, 4-10, 13-25, 52, 99 and 101 or a complement thereof under moderately stringent conditions.
- 4. A polypeptide comprising an immunogenic portion of a *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 26-51, 138, 139, 163-183, 201, 240, 242-247, 253-256, 295-298, 309, 316, 318-320, 322, 324, 328, 329, 333, 335, 337, 339 and 341, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 26-51, 138, 139, 163-183, 201, 240, 242-247, 253-256, 295-298, 309, 316, 318-320, 322, 324, 328, 329, 333, 335, 337, 339 and 341 or a complement thereof under moderately stringent conditions.
- 5. A DNA molecule comprising a nucleotide sequence encoding a polypeptide according to any one of claims 1-4.
- 6. An expression vector comprising a DNA molecule according to claim 5.

- 7. A host cell transformed with an expression vector according to claim 6.
- 8. The host cell of claim 7 wherein the host cell is selected from the group consisting of *E. coli*, yeast and mammalian cells.
- 9. A pharmaceutical composition comprising one or more polypeptides according to any one of claims 1-4 and a physiologically acceptable carrier.
- 10. A pharmaceutical composition comprising one or more DNA molecules according to claim 5 and a physiologically acceptable carrier.
- 11. A pharmaceutical composition comprising one or more DNA sequences recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and a physiologically acceptable carrier.
- 12. A vaccine comprising one or more polypeptides according to any one of claims 1-4 and a non-specific immune response enhancer.
  - 13. A vaccine comprising:
- a polypeptide having an N-terminal sequence selected from the group consisting of sequences recited in SEQ ID NO: 134 and 135; and
  - a non-specific immune response enhancer.

#### 14. A vaccine comprising:

one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID Nos.: 3. 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and

a non-specific immune response enhancer.

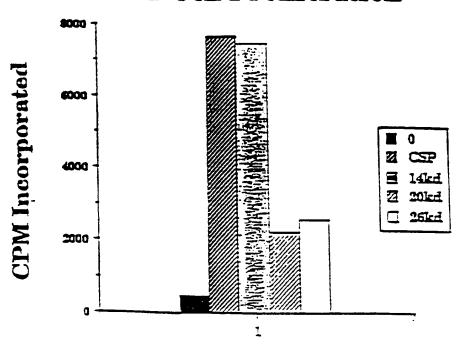
- 15. The vaccine of claims 12-14 wherein the non-specific immune response enhancer is an adjuvant.
- 16. A vaccine comprising one or more DNA molecules according to claim 5 and a non-specific immune response enhancer.
- 17. A vaccine comprising one or more DNA sequences recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and a non-specific immune response enhancer.
- 18. The vaccine of claims 16 or 17 wherein the non-specific immune response enhancer is an adjuvant.
- 19. A method for inducing protective immunity in a patient, comprising administering to a patient a pharmaceutical composition according to any one of claims 9-11.
- 20. A method for inducing protective immunity in a patient, comprising administering to a patient a vaccine according to any one of claims 12-18.
- 21. A fusion protein comprising two or more polypeptides according to any one of claims 1-4.
- 22. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and ESAT-6.
- 23. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and the *M. tuberculosis* antigen 38 kD (SEQ ID NO:155).

- 24. A pharmaceutical composition comprising a fusion protein according to any one of claims 21-23 and a physiologically acceptable carrier.
- 25. A vaccine comprising a fusion protein according to any one of claims 21-23 and a non-specific immune response enhancer.
- 26. The vaccine of claim 25 wherein the non-specific immune response enhancer is an adjuvant.
- 27. A method for inducing protective immunity in a patient, comprising administering to a patient a pharmaceutical composition according to claim 24.
- 28. A method for inducing protective immunity in a patient, comprising administering to a patient a vaccine according to claims 25 or 26.
  - 29. A method for detecting tuberculosis in a patient, comprising:
- (a) contacting dermal cells of a patient with one or more polypeptides according to any one of claims 1-4; and
- (b) detecting an immune response on the patient's skin and therefrom detecting tuberculosis in the patient.
  - 30. A method for detecting tuberculosis in a patient, comprising:
- (a) contacting dermal cells of a patient with a polypeptide having an N-terminal sequence selected from the group consisting of sequences recited in SEQ ID NO: 134 and 135; and
- (b) detecting an immune response on the patient's skin and therefrom detecting tuberculosis in the patient.
  - 31. A method for detecting tuberculosis in a patient, comprising:
- (a) contacting dermal cells of a patient with one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308,

- 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and
- (b) detecting an immune response on the patient's skin and therefrom detecting tuberculosis in the patient.
- 32. The method of any one of claims 29-31 wherein the immune response is induration.
  - 33. A diagnostic kit comprising:
  - (a) a polypeptide according to any one of claims 1-4; and
- (b) apparatus sufficient to contact said polypeptide with the dermal cells of a patient.
  - 34. A diagnostic kit comprising:
- (a) a polypeptide having an N-terminal sequence selected from the group consisting of sequences recited in SEQ ID NO: 134 and 135; and
- (b) apparatus sufficient to contact said polypeptide with the dermal cells of a patient.
  - 35. A diagnostic kit comprising:
- (a) a polypeptide encoded by a DNA sequence selected from the group consisting of SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and
- (b) apparatus sufficient to contact said polypeptide with the dermal cells of a patient.

- 36. A diagnostic kit comprising:
- (a) a fusion protein according to any one of claims 21-23; and
- (b) apparatus sufficient to contact said fusion protein with the dermal cells of a patient.
- 37. A fusion protein according to claim 23 comprising an amino acid sequence selected from the group consisting of sequences recited in SEQ ID NO: 153, 209, 351 and 355.

## D7 T Cell Proliferation



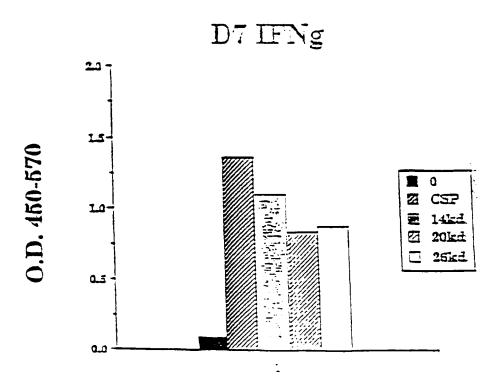
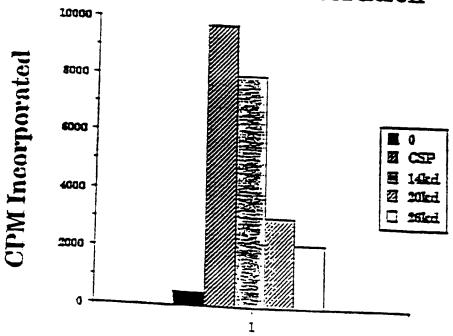


FIG. : =

# D160 T Cell Proliferation



## D160 IFNg

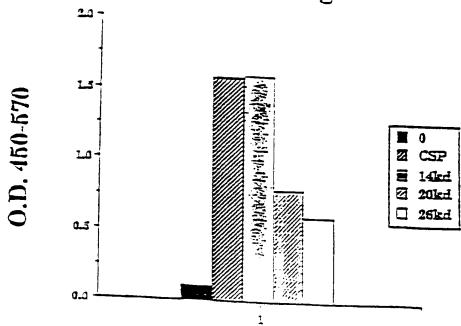
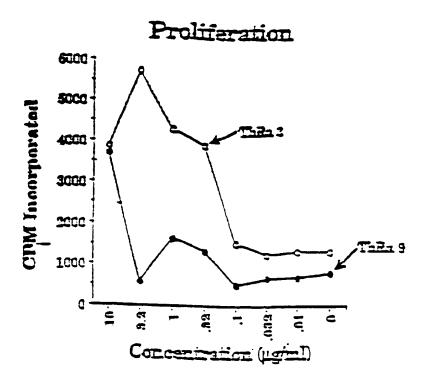


FIG. 1B



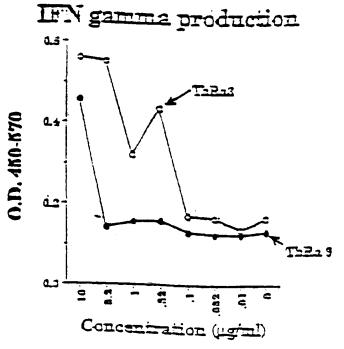
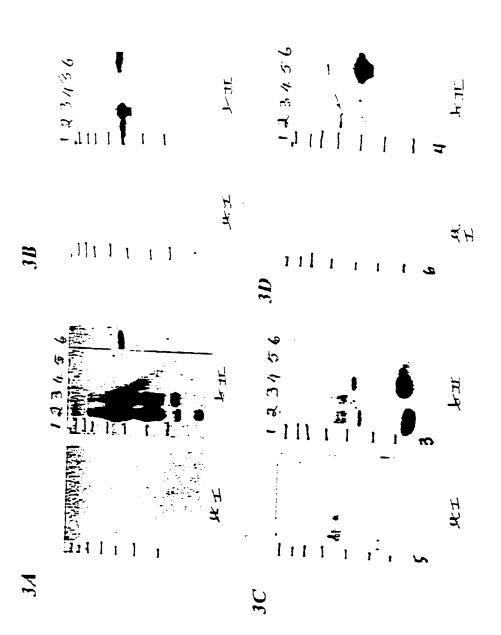
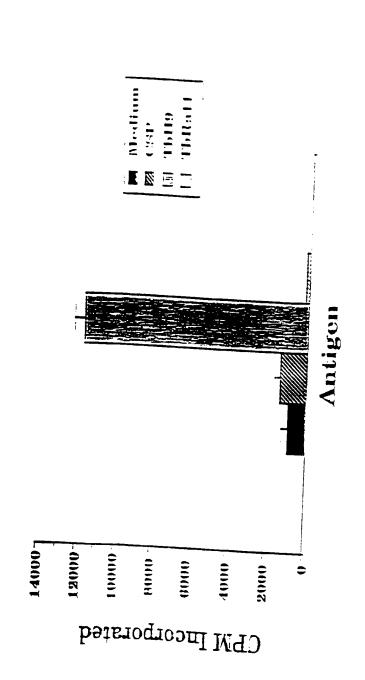


FIG. 2



FIGS. 3 A-D

# responds poorly to CSF T cell clone 131TbH9



F16. 4/

T Cell Clone PPD 800-10 IFNg Production

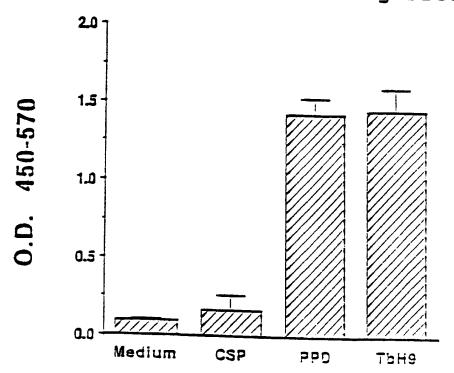
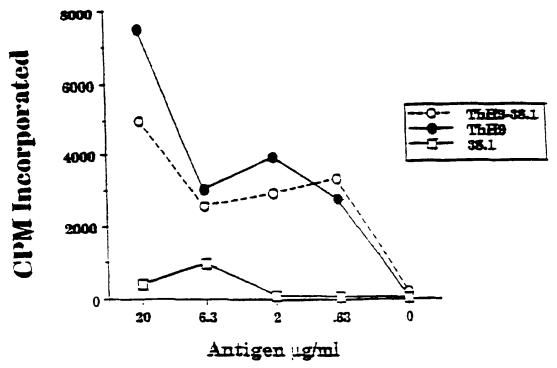
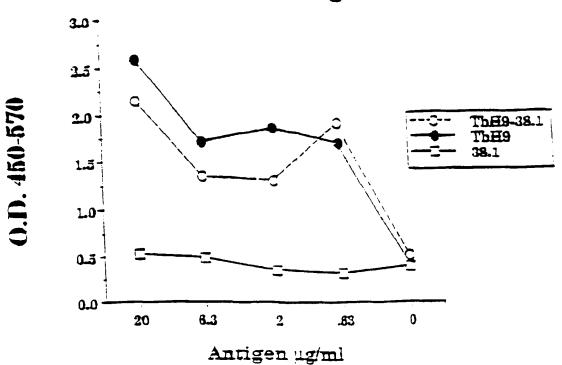


FIG. 4B

### D131 T Cell Proliferation

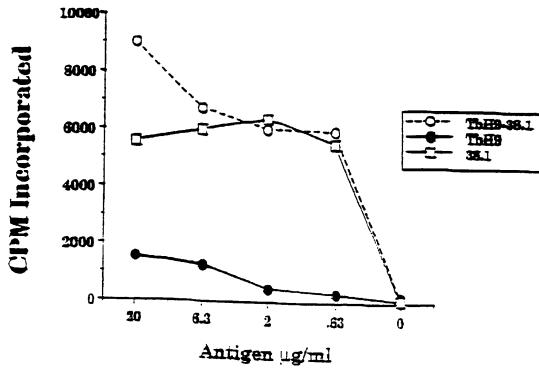


D131 IFNg

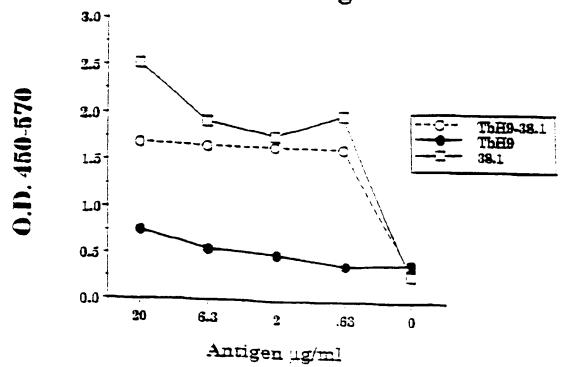


FIGS. 5.4-B

## D184 T Cell Proliferation

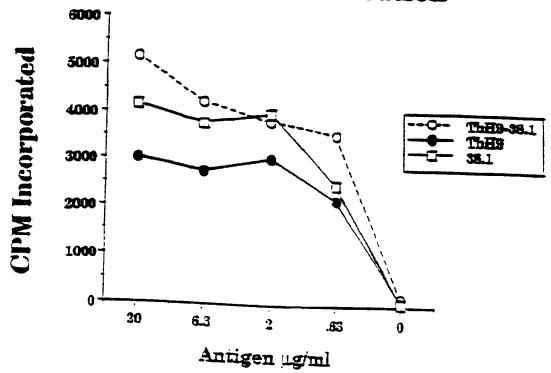


D184 IFNg

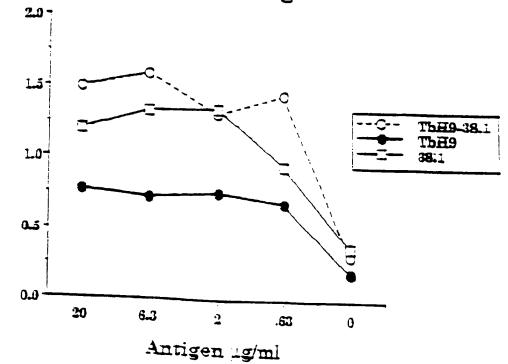


FIGS. 6.4-B

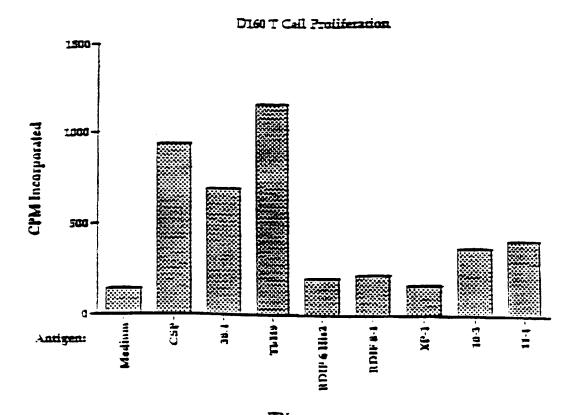
# D201 T Cell Proliferation

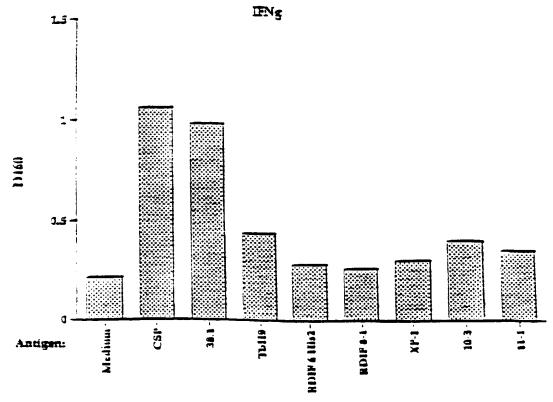


D201 IFNg

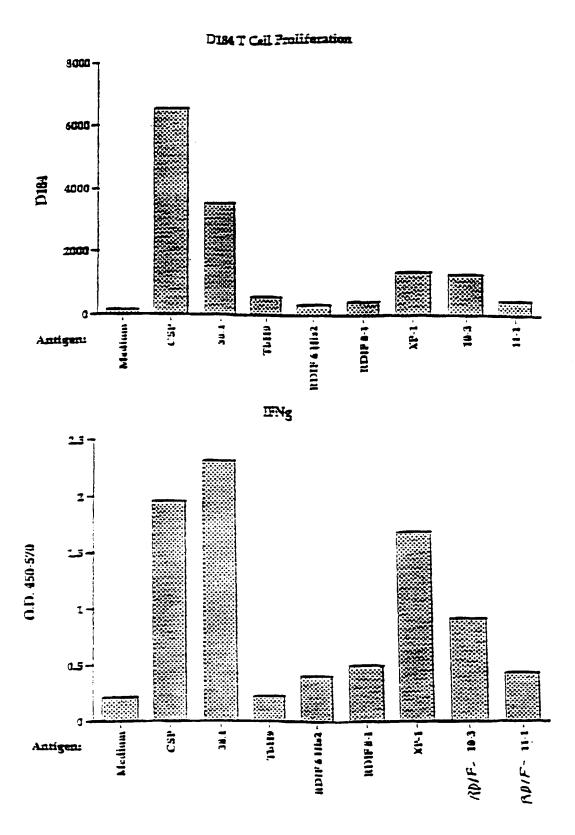


FIGS. 7.4-B



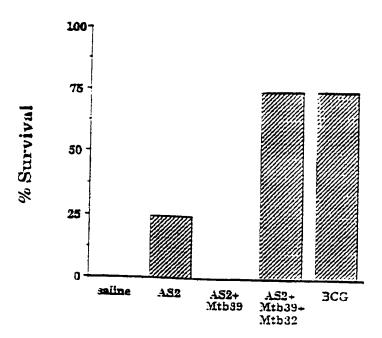


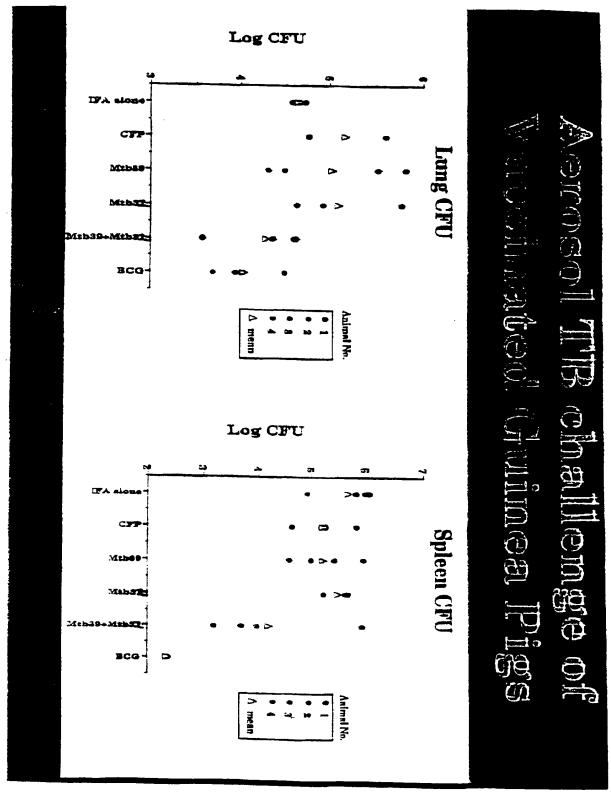
FIGS. 8A-B



FIGS. 9A-B

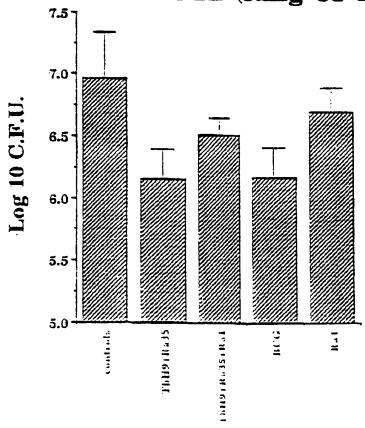
# Tuberculosis: Protection of Cynomolgus Monkeys with Recombinant Antigens of Mtb





FIGS. 11 A-B

# DNA Immunized mice challenged with aerosol TB (lung CFU)



### **PCT**

### WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



### INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

ļ	(51) International Patent Classification 6:		(11) International Publication Number:	WO 99/42076
	A61K 39/04, 49/00, G01N 33/53, C07K	A3		
	1/00, C12N 15/00		(43) International Publication Date:	26 August 1999 (26.08.99)

(21) International Application Number: PCT/US99/03268

(22) International Filing Date: 17 February 1999 (17.02.99)

(30) Priority Data:

09/025,197 18 February 1998 (18.02.98) US 09/072,967 5 May 1998 (05.05.98) US

(71) Applicant: CORIXA CORPORATION [US/US]; 1124 Columbia Street, Seattle, WA 98104 (US).

(72) Inventors: REED, Steven, G.; 2843 - 122nd Place, N.E., Bellevue, WA 98005 (US). SKEIKY, Yasir, A., W.; 8327 25th Street, Seattle, WA 98107 (US). DILLON, Davin, C.; 21607 N.E. 24th Street, Redmond, WA 98053 (US). CAMPOS-NETO, Antonio; 9308 Midship Court, N.E., Bainbridge Island, WA 98110 (US). HOUGHTON, Raymond; 2636 - 242nd Place, S.E., Bothell, WA 98021 (US). VEDVICK, Thomas, S.; 124 South 300th Place, Federal Way, WA 98003 (US). TWARDZIK, Daniel, R.; 10195 South Beach Drive, Bainbridge Island, WA 98110 (US). LODES, Michael, J.; 9223 - 36th Avenue S.W., Seattle, WA 98126 (US). HENDRICKSON, Ronald, C.; 4114 S.W. Charlestown Street, Seattle, WA 98116 (US).

(74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

### Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report: 14 October 1999 (14.10.99)

(54) Title: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS

### (57) Abstract

Compounds and methods for inducing protective immunity against tuberculosis are disclosed. The compounds provided include polypeptides that contain at least one immunogenic portion of one or more *M. tuberculosis* proteins and DNA molecules encoding such polypeptides. Such compounds may be formulated into vaccines and/or pharmaceutical compositions for immunization against *M. tuberculosis* infection, or may be used for the diagnosis of tuberculosis.

### FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Сапада	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	ΚZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

### INTERNATIONAL SEARCH REPORT

International application No. PCT/US99/03268

A. CLASSIFICATION OF SUBJECT MATTER				
IPC(6) : A61K 39/04, 49/00; G01N 33/53; C07K 1/00; C12N 15/00				
US CL	: 424/9.2, 248.1; 435/7, 172.3, 320, 320.7; 530/350	and the second second		
According t	to International Patent Classification (IPC) or to both	national classification and IPC		
B. FIEL	.DS SEARCHED			
Minimum d	ocumentation searched (classification system followed	d by classification symbols)		
U.S. :	424/9.2, 248.1; 435/7, 172.3, 320, 320.7; 530/350			
Documentat	tion searched other than minimum documentation to the	extent that such documents are included	in the fields searched	
			annah tarma urad)	
Electronic d	lata base consulted during the international search (na	ime of data base and, where practicable,	scarch lettins used)	
Please Se	e Extra Sheet.			
<del></del>				
C. DOC	UMENTS CONSIDERED TO BE RELEVANT			
Category*	Citation of document, with indication, where ap	propriate, of the relevant passages	Relevant to claim No.	
v	OFTINCED T at al. Claning an	d B cell enitone manning of	1, 5-8, 10, 11,	
X	OETTINGER, T. et al. Cloning an		13, 14, 33	
	MPT64 from Mycobacterium tubercu		15, 14, 55	
Y	Immunity. May 1994, Vol. 62, No. 5	, pages 2058-2064, especially		
	Figure 1.		2-4, 9, 12, 15-32	
Y	US 5,108,745 A (HORWITZ) 28 Apr	il 1992, column 1, line 25 to	12-20, 27-28	
-	column 13, line 54.			
	20, 1			
		[		
		ļ		
V Furth	her documents are listed in the continuation of Box C	. See patent family annex.		
<u> </u>				
·	ecial categories of cited documents:	*T* later document published after the inte date and not in conflict with the appl	ication but cited to understand	
	cument defining the general state of the art which is not considered be of particular relevance.	the principle or theory underlying the	invention	
*E* car	rlier document published on or after the international filing date	"X" document of particular relevance; the considered novel or cannot be consider		
	cument which may throw doubts on priority claim(s) or which is	when the document is taken alone	ed w airone ai aironaro sup	
cit	ed to establish the publication date of another estation or other scial reason (as specified)	"Y" document of particular relevance; the	s claimed invention cannot be	
•		considered to involve an inventive combined with one or more other such		
	cument referring to an oral disclosure, use, exhibition or other pans	being obvious to a person skilled in t		
	cument published prior to the international filing date but later than	*&* document member of the same patent	family	
	actual completion of the international search	Date of mailing of the international sea	rch report	
	1 8 AUG 1999			
09 JULY	1999	10 700 1000		
N .	- IV ddagg of the ISA/IIS	Authorized officer		
Name and r	mailing address of the ISA/US iner of Patents and Trademarks	Admonized officer	100	
Box PCT	Box PCT RODNEY P. SWARTZ, PH.D. // (\N )		1 (iV)	
	n, D.C. 20231	,		
Facsimile No. (703) 305-3230 Tel		Telephone No. (703) 308-0196		

### INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/03268

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
X  Y	PAL, P.G. et al. Immunization with extracellular proteins of <i>Mycobacterium tuberculosis</i> induces cell-mediated immune responses and substantial protective immunity in a guinea pig model of pulmonary tuberculosis. Infection and Immunity. November 1992, Vol. 60, No. 11, pages 4781-4792, especially 4782.	1-4, 9, 12-15, 19- 20 21-28
Y	US 4,689,397 A (SHINNICK et al) 25 August 1987, column 1, line 15 to column 22, line 6.	6-8, 21-37
X  Y	US 5,714,593 A (LAQUEYRERIE et al) 03 February 1998, entire document.	1-26  27-37

### INTERNATIONAL SEARCH REPORT

International application No. PCT/US99/03268

B. FIELDS SEARCHED Electronic data bases consulted (Name of data base and where practicable terms used):		
APS, BIOSIS, CABA, CAPLUS, EMBASE, LIFESCI, MEDLINE, SCISEARCH, MASPAR search terms: tuberculosis, antigen, extracellular, soluble, protein, polypeptide, vaccine, polynucleotide, nucleic acid, amino acid, specific sequence identification numbers.		